

# U. OF F. LIBRARIES

414  
K57P  
1971  
C.2

UNIVERSITY  
OF FLORIDA  
LIBRARIES



PHONOLOGICAL CHANGE

Paul Kiparsky

Reproduced by the  
Indiana University  
Linguistics Club  
November, 1971

U. OF F. LIBRARIES

OVERSIZE

414  
K57  
1971



22010001330000

### Abstract

There are two processes that bring about change in the phonological components of languages: sound change and imperfect learning. Sound change has since H. Paul been correctly regarded as a composite product of change in the way phonological representations are realized and resulting restructuring in the phonological representations themselves. According to the neo-grammatician view, shared by most varieties of structuralism, changes in the realization of phonological representations are due to deviations in performance (*parole*), whereas the theory of generative grammar instead suggests that they are due to changes in competence (*langue*), namely to the adoption of new phonological rules by speakers. The latter view is supported both by theoretical arguments and by the empirical evidence of actual sound changes that depend in various ways on non-phonetic representations and on the abstract form of phonological rules.

Imperfect learning is subject to the constraints that it must not increase the complexity of a grammar, and furthermore that it must result in only minor modifications of the output. There are several forms of imperfect learning, the most interesting of them being reordering of rules, with a specific directionality that is apparently due to the greater simplicity of one type of order.

Many traditional results of historical linguistics are in need of reexamination, particularly as regards the relative chronology of prehistoric sound changes. For example, Grassmann's Law is not an independent innovation in Greek and Indic, as has been universally assumed, but rather a joint innovation which, moreover, apparently antedates the split of the two dialects.

## TABLE OF CONTENTS

	Page No.
PREFACE	iii
CHAPTER 1. SOUND CHANGE	
1.1 The Neo-grammarian Theory	1
1.2 Restructuring	3
1.3 Innovation	8
1.4 Phonetic Leaps	11
1.5 Non-phonetic Environments	17
1.6 The Placement of Innovations	22
CHAPTER 2. IMPERFECT LEARNING	
2.1 The Problem	25
2.2 Loss and Reordering	30
2.3 Unmarked Order	39
2.4 Proportional Analogy	45
2.5 Further Types	52
CHAPTER 3. RECONSTRUCTION	
3.1 Relative Chronology	57
3.2 Grassmann's Law: Synchronic Description	60
3.3 Grassmann's Law: Internal Reconstruction	66
3.4 Grassmann's Law: Comparative Reconstruction	69
FOOTNOTES	77
BIBLIOGRAPHY	85

### Preface

I discuss here some questions of diachronic phonology in terms of the general approach to linguistic diachrony developed by Halle, Matthews and Klima and applied by them and many others to the study of empirical problems in both phonology and syntax. The emphasis in Chapters 1 and 2 is on implications for linguistic theory; Chapter 3 then takes up problems of internal and comparative linguistic reconstruction, particularly as regards relative chronology of phonological changes. I try throughout to point out the similarities and differences between the conclusions reached here and more traditional views.

I have tried to pick as familiar examples as possible, taking more obscure ones only when they have the compensating advantage of simplicity. Most of the examples have been drawn from Germanic, the field with which I am most familiar, but several other languages are also represented. The most extended discussion, in fact, is devoted to Grassmann's Law and its relation to various other well-known rules of Greek and Indic phonology.

The notation employed here is the standard one. When necessary, I use square brackets [ ] to enclose phonetic representations and slants / / to enclose non-phonetic forms, which need not necessarily be underlying forms but may be intermediate stages of representation. When the status of forms I am discussing is obvious, unknown, or irrelevant, I use underlining instead. I use an arrowhead to link original changed forms, e.g. dhadhami > dadhāmi. Incorrect forms are enclosed in quotation marks to distinguish them from reconstructed forms, which are affixed with the usual asterisk.

The individual chapters of this thesis have benefited from many people. With regard to Chapter 1, I would like to thank P. M. Postal for puncturing a false theory of mine with an appropriate counter-example, and to G. Reichenkron for guidance in questions of Romance Philology. Chapter 2 has been influenced by discussions with B. C. Hall, Z. S. Harris and R. J. Stanley, and especially with T. G. Bever and G. H. Matthews. Some of this material was presented at the Graduate Linguistics Circle of Columbia University, where discussions with E. Garcia and others helped me to see things more clearly, and also at U.C.L.A. and at the Winter, 1964 meeting of the Linguistic Society of America.

But especially I would like to thank my adviser, M. Halle, with whom I have argued at length over most of these problems and who has

been a constant source of ideas and encouragement, also suggesting many improvements in the presentation of this material. In a more indirect but equally important way this thesis also owes much to N. Chomsky and R. Jakobson.

## Chapter I

### Sound Change

#### 1.1 The Neo-grammatician Theory

The clearest and most influential formulation of the neo-grammatician theory of phonological change is found in the third chapter of Paul's Grundzüge der Sprachgeschichte. According to Paul, phonological information is stored by the speaker in two forms: as auditory representations (Lautbilder) and as articulatory representations (Bewegungsgefühle). One form of phonological change (Lautwandel) takes place in that the execution of articulatory movements changes and that these changes in turn lead to changes in the articulatory representations themselves. There are two ways in which the execution of articulatory movements can change and accordingly two major manifestations of Lautwandel: articulatory drift (Verschiebung des Bewegungsgefühles) and replacement (Vertauschung der Elemente).

In articulatory drift, the articulation of the minimal phonetic elements (it is not clear what these elements are) undergoes gradual shift. The shift is subject to the constraint that the resulting auditory deviation must not transcend the threshold of conscious perception (Kontrolle des Lautbildes). Phonological change takes place in that the articulatory representations are continually revised to match these shifts in their execution. Aware that merely random fluctuations of articulation could have no effect, Paul proposes that a tendency towards greater facility of articulation lends a directionality to the drift. But what facility of articulation is and how it is to be determined is left unspecified except for the statement that assimilation increases it, and that it depends in part on the phonological systems of the language. It has been long realized that this gap in Paul's theory is impossible to fill, there being no way of defining facility of articulation so that it would account for what sound changes actually happen. Hence the idea of articulatory drift is in effect quite empty.

In replacement, phonetic elements are replaced by others in individual productions of words because of slips of the tongue. These slips of the tongue become part of the language through being adopted as correct forms by oncoming generations of language learners. As examples of replacement Paul mentions metathesis, assimilation and dissimilation, and presumably epenthesis and apocopation should be added to complete his list.

But there is another way, according to Paul, in which phonological changes amounting to the same effect, largely, as articulatory drift and replacement may take place. This is that children deviate from their linguistic models and from the beginning form articulatory representation which differ from those of the speakers who provide their linguistic experience. Changes of this type must be sufficiently insignificant to escape correction; they can thus either result in the kinds of metatheses, assimilations, etc. brought about by replacement or else in differences of articulatory representation with minimal acoustic consequences, such as that of dorsal and alveolar t [sic].

In this respect, there is a curious equivocation in Paul as to which of the postulated mechanisms is to be considered the basic form of phonological change. On the one hand, he very clearly points out that articulatory drift in mature speakers cannot lead to changes of any consequence:

"Innerhalb der nämlichen Generation werden auf diese Weise (sc. by articulatory drift) immer nur sehr geringfügige Verschiebungen zu Stande kommen. Merlichere Verschiebungen erfolgen erst, wenn eine ältere Generation durch eine neu heranwachsende verdrängt ist." (62-63)

He emphasizes that it is by imperfect learning that phonological change in practice most commonly takes place:

"Man wird also sagen können, dass die Hauptveranlassung zum Lautwandel in der Übertragung der Laute auf neue Individuen liegt. (His emphasis). Für diesen Vorgang ist also der Ausdruck Wandel, wenn man sich an das wirklich tatsächliche hält, gar nicht zutreffend, es ist vielmehr eine abweichende Neuerzeugung." (63)

On the other hand, the famous principle of regularity is claimed only for articulatory drift in mature speakers (§47-8), i.e. for a form of phonological change which in the above quotations is dismissed as insignificant. However, when discussing regularity Paul contradicts his earlier statements by implying that articulatory drift in mature speakers is on the contrary the major form of phonological change:

"Es bleiben nun allerdings einige Arten von lautlichen Veränderungen übrig, für die sich konsequente Durchführung theoretisch nicht als notwendig erweisen lässt. Diese bilden aber eine verhältnismässig geringen Teil der gesamten Lautveränderungen, und sie lassen sich genau abgrenzen. Einerseits also gehören hierher die Fälle, in denen ein Laut vermittelst einer abweichenden Artikulation nachgeahmt wird, anderseits die §45 besprochenen Metathesen, Assimilationen und Dissimilationen." (73)

As more important than the inadequacies and inconsistencies of Paul's theory I would regard his fundamental insight that the phonology of a language can change in two ways: by Lautwandel and by abweichende Neuerzeugung. I will translate them as sound change and imperfect learning, respectively, subsuming both under the general term phonological change. Sound change is brought about by the joint operation of innovation and subsequent restructuring. Innovation in phonology is change in the way phonological representations are executed by speakers of a language; restructuring is the resulting revision in the phonological representations. As the latter were considered to be phonetic by Paul, and restructuring immediately followed on innovation, the distinction between innovation and restructuring was to him purely theoretical and devoid of any practical consequences. But with richer conceptions of phonological structure it takes on a cardinal importance for diachronic phonology.<sup>1</sup> Imperfect learning is due to the fact that the child does not learn a grammar directly but must recreate it for himself on the basis of a necessarily limited and fragmentary experience with speech. It is in no way surprising that the grammar should change in the process of transmission across generations of speakers.

In the following discussion of phonological change I accept these basic concepts, though interpreting them (particularly the concept of imperfect learning) very differently from Paul. This chapter deals with restructuring and innovation, the presentation taking the form of a confrontation of the taxonomic and generative phonological views of these processes. Chapter 2 then takes up imperfect learning, an aspect of phonological change whose importance has, I think, been grossly underrated.

## 1.2 Restructuring

Take the case of Old High German umlaut, a canonical example of diachronic phonemics.<sup>2</sup> In Old High German, vowels became non-grave before i; short vowels were in addition made non-compact. Subsequently unstressed vowels were reduced to e (probably phonetically a). Cf., with change of gravity,

- |                           |                                     |
|---------------------------|-------------------------------------|
| (1) wurmi > würmi > würme | (Nom. Pl. of <i>wurm</i> 'worm')    |
| tāti > tāti > tāte        | (Nom. Pl. of <i>tat</i> 'deed')     |
| nōti > nōti > nōte        | (Nom. Pl. of <i>nōt</i> 'distress') |

and with change of both gravity and compactness,

- |                           |                                    |
|---------------------------|------------------------------------|
| (2) slagi > slegi > slege | (Nom. Pl. of <i>slag</i> 'stroke') |
| gasti > gesti > geste     | (Nom. Pl. of <i>gast</i> 'guest')  |

What is the effect of these two sound changes on the grammar?

After umlaut, taxonomic phonemics requires a restructuring of phonetic representations in the case of just one of the vowels affected by the change. Changes like (1) involve merely the introduction of new allophones before i, but the change of a to e as in (2) must be regarded as a phonemic change because a and e contrast in other environments. Thus a taxonomic grammar must at this stage phonemically represent nōti as /nōti/ but slegi as /slegi/ and must, if it is to account for the regularities of the language, contain two separate umlaut rules, a morphophonemic umlaut rule turning a into e (e.g. /slagi/ > /slegi/) to provide the phonemic level of representation, and a phonemic umlaut rule affecting the remaining vowels (e.g. /nōti/ > [nōti]) to provide the phonetic level of representation. Vowel reduction then creates minimal pairs between each umlaut vowel and its underlying counterpart, cf. Nom. Pl. würme with Dat. Sg. wurme. With this second innovation all umlaut vowels become phonemic and umlaut is henceforth entirely a "non-automatic" morphophonemic alternation.

The theory of generative grammar suggests a radically different picture. Since a grammar is a system of rules, it becomes natural to regard innovations as added rules (Halle, 1962; Klima, 1964). Thus, umlaut is the addition of the following phonological rule to the grammar of Old High German:

$$(3) \quad \left[ \begin{array}{l} \text{[- consonantal]} \\ \text{[- long]} \end{array} \right] \rightarrow \left[ \begin{array}{l} \text{[- grave]} \\ \text{[- compact]} \end{array} \right] / \underline{\underline{C_n}} \quad i$$

(Material in angled brackets must be either all present or all absent in applying rules;  $C_n$  denotes "n or more consonants".)

Once we abandon proceduralism and with it the level of taxonomic phonemics there no longer remains the slightest reason to suppose that the phonemic representations of Old High German were revised as a result of (3), and that the grammar came to contain two separate umlaut rules. The grammar of this stage, in fact, has the same phonemic representations as the pre-umlaut stage (e.g. /slagi, nōti/) and contains the phonological rule (3) which derives the correct phonetic forms (e.g. slegi, nōti). Nor does the subsequent development of minimal pairs like würme : wurme by the addition of the vowel reduction rule

$$(4) \quad \left[ \begin{array}{l} \text{[+ vocalic]} \\ \text{- consonantal} \\ \text{- stressed} \end{array} \right] \rightarrow [ə]$$

necessarily mean that the umlaut vowels have become phonemic and that the occurrence of umlaut is no longer predictable. There is the relevant fact that any grammar will at that stage still have to contain

both rules (3) and (4) for reasons that are quite independent of such forms as (1) and (2). (3) is required in the grammar because umlaut is still phonologically conditioned by i in those suffixes that were protected from (4) by a secondary stress. Cf., with umlaut before i, tōrlisch 'foolish' (tōre 'fool'), bēstīn 'made of bast' (bast 'bast'), gēstīnne 'female guest' (gast 'guest'), tüechlin 'cloth (Dim.)' (tuoch 'cloth'); but, without umlaut, mānūnge 'exhortation' (manen 'exhort'), ármuot 'poverty' (arm 'poor'). And (4) is a phonological rule because stress alternations such as those between prefix and stem are accompanied by automatic shifts in vowel quality, e.g. ántlaz 'absolution', entlāzen 'absolve'; ímbiz 'partaking', embizen 'partake'; úrteil 'judgment', ertéilen 'judge'. Given the necessity of including (3) and (4) in the grammar, it is now only necessary to make two further assumptions in order to account for forms like (1) and (2): that the Plural ending in these forms is phonemically i, and that the rules apply in the order (3, 4). This would in fact seem to be the correct solution for Middle High German, though no longer for Modern German.

In this way the more highly structured form of language postulated by the theory of generative grammar, together with the goal of giving in a grammar an explicit account of the regularities of a language, often leads to a synchronically motivated retention of earlier underlying representations and incorporation of innovations as phonological rules, often in a synchronic order that matches their relative chronology, in situations where a taxonomic account would require immediate restructuring. By the same token, it follows from generative grammatical theory that restructurings, when they do take place, can be of a far more radical nature than taxonomic phonemics would have one suspect. Clearly, too, it depends on the structure of the language whether an innovation leads to restructuring. Other things being equal, an innovation is more likely to remain as a synchronic rule in a language with a complicated phonology than in one with very few rules. Thus complicated phonologies tend to remain complicated and simple phonologies tend to remain simple: hence the relative stability of typological properties such as agglutination, inflection and isolation.

The sequence of sound changes known as Grimm's Law furnishes a good illustration of restructuring.<sup>3</sup> For a stage of Pre-Germanic preceding the operation of Grimm's Law we can assume the existence of the verb stems skab 'shape', nas 'save', with infinitive and part participle forms

(5) skabian      skabtas      nasian      nasitas

phonetically realized as

(6) skabjan      skaptas      nasjan      nasitas.

The devoicing of b in skaptas is due to a rule of regressive assimilation of Indo-European origin:

(7) [+ obstruent] → [α voiced] / \_\_\_\_\_ { + obstruent  
α voiced

The vowel i automatically becomes a glide before vowels by a rule that I will tacitly assume to apply in the following.

The first step of Grimm's Law is that voiceless stops are aspirated (made tense) everywhere except after obstruents:

(8)	<table border="1"> <tr> <td>- continuant</td><td></td></tr> <tr> <td>+ obstruent</td><td></td></tr> <tr> <td>- voiced</td><td></td></tr> </table>	- continuant		+ obstruent		- voiced		$\rightarrow$ [+ tense] / { [- obstruent] # }
- continuant								
+ obstruent								
- voiced								

Rule (8) changes skaptas to skap<sup>h</sup>tas and nasitas to nasithas, thus introducing an alternation tas : thas in the past participle morpheme. It does not lead to restructuring; the result of adding (8) is a grammar with the underlying representations (5) and the rules (7) and (8), in that order, yielding the phonetic forms

(9) skabjan skaphtas nasjan nasithas.

By the next step of Grimm's Law, aspirated stops turn into continuants:

(10)  $\left[ \begin{array}{l} +\text{tense} \\ +\text{obstruent} \end{array} \right] \rightarrow [+ \text{continuant}]$

This rule applies both to the voiceless aspirates which arose by (8), e.g. skap<sup>h</sup>tas > skaftas and nasith<sup>h</sup>as > nasiθas, and to the voiced aspirates inherited from Indo-European, e.g. bheran > þeran (cf. Sanskrit bharati, Engl. bear). The alternation of aspiration now appears as a stop - continuant alternation. One possible grammar for this stage of course has the same dictionary representations as before, with the rules (7, 8, 10). But at this point there is a quite different grammar which accounts for the same language in a simpler way. The old voiced aspirates bh etc. are replaced in the dictionary representations by the corresponding continuants þ etc. The restructured grammar has the rule (7), by which skabtas > skaptas, followed by a new rule (11) that combines the effect of (8) and (10), turning skaptas > skaftas and nasitas > nasiθas.

$$(11) \left[ \begin{array}{l} +\text{obstruent} \\ -\text{voiced} \end{array} \right] \rightarrow [+\text{continuant}] / \left\{ \begin{array}{l} [-\text{obstruent}] \\ \# \end{array} \right\} \quad \underline{\hspace{2cm}}$$

With this restructuring the feature of tenseness has ceased to play a role in the obstruent system, and has been replaced by the feature of continuance. Thus the very simple rule (10) has triggered considerable revision in both the rules and underlying representations of the language.

The final part of Grimm's Law makes the voiced stops unvoiced (*skabjan* > *skapjan*):

$$(12) \left[ \begin{array}{l} +\text{obstruent} \\ -\text{continuant} \end{array} \right] \rightarrow [-\text{voiced}]$$

This gives the phonetic forms

$$(13) \text{ skapjan } \text{ skaftas } \text{ nasjan } \text{ nasiθas}$$

Again a possible grammar for this stage would retain the same phonemic system with the phonological rules (7, 11, 12). But there is a simpler grammar in which the old voiced stops *b* etc. are replaced by the corresponding voiceless stops *p* etc. in the dictionary entries; and the old voiceless stops *p* etc. are replaced by the corresponding continuants *f* etc., so that the underlying forms become

$$(14) \text{ skapian } \text{ skapθas } \text{ nasian } \text{ nasiθas}$$

The phonetic forms (13) are derived from (14) by rule (15), a kind of anti-Grimm's Law:

$$(15) [+ \text{obstruent}] \left[ \begin{array}{l} +\text{obstruent} \\ -\text{strident} \end{array} \right] \rightarrow [+\text{continuant}] . [-\text{continuant}]$$

(Rule (7) remains in the language for other reasons.)

It is facts of this kind that constitute the ultimate linguistic justification for the distinction between innovation and restructuring. To attain the level of descriptive adequacy, a theory of linguistic change must express the limitations that obtain on the possible differences of successive grammars of a language, which are very severe indeed. It would be totally impossible to do this under the assumption that successive optimal grammars of a language somehow directly change into each other, e.g. by meta-rules that map grammars directly into their successors. The significant regularities do not lie in any direct relation between successive grammars but rather in the form of the innovations that bring the changes about. Cf. Saussure's remark that "there is no inner bond between the initial fact and the effect that it may subsequently produce on the whole system" (1959, 87).

This is also what invalidates P. Garde's assertion: "ou bien les "règles synchroniques" sont une nouvelle formulation des lois diachroniques, ou bien elles sont fausses." (1965). The synchronic rules of a grammar not only need not be reformulations of past innovations but need not even bear the slightest resemblance to them, as the above examples show. A corollary and equally obvious moral to be drawn from restructuring is that there is no point in trying to do internal reconstruction by mindlessly translating the rules of a grammar into historical statements (or conversely trying to write generative grammars by providing the arrowheads of historical grammars with shifts). Explicit linguistic descriptions are a prerequisite but not a substitute for linguistic reconstruction.

### 1.3 Innovation

There is one central difference which above all marks off Halle's conception of linguistic innovation from virtually all of neo-grammarians and American structuralist thinking on the subject. This difference is that in Halle's view linguistic innovations, since they are added grammatical rules, originate in langue, in direct contradiction to the view which is quite explicit in Saussure and implicit in neo-grammarians and American structuralist theory that linguistic innovations originate in parole. In the linguistic theory of Saussure, the idea that change in language originates in parole is internally well motivated by two specific aspects of the way in which he conceived of the difference between langue and parole. The first of these is Saussure's view of langue as a social phenomenon in contrast to parole as an individual phenomenon. Since innovations begin with individual speakers and are only later adopted by the whole speech community, it then follows that innovations are part of parole before they become part of langue:

"Once in possession of this double principle of classification, we can add that everything diachronic in language is diachronic only by virtue of speaking. It is in speaking that the germ of all change is found. Each change is launched by a certain number of individuals before it is accepted for general use." (98)

The second, which pertains specifically to sound change, is Saussure's exclusion of phonetics from the domain of langue:

"An argument against separating phonation from language might be phonetic changes, the alterations of the sounds which occur in speaking and which exert such a profound influence on the future of language itself. Do we really have the right to pretend that language exists independently of phonetic change?"

Yes, for they affect only the material substance of words. If they attack language as a system of signs, it is only indirectly, through subsequent changes of interpretations; there is nothing phonetic in the phenomenon." (18)

Thus Saussure regards sound change as an ultimately extra-linguistic phenomenon. It can hence consist only of the kinds of phenomena that may conceivably be brought about by extraneous deviations of the speech mechanism, i.e. precisely slips of the tongue and imperceptible articulatory drifts. Thus Saussure has from a rather different theoretical basis arrived at essentially the neo-grammatician constraints on sound change.

But the view which ascribed innovation to parole is only as sound as the presuppositions on which it is based. Both the idea of langue and parole as social and individual phenomena respectively and the exclusion of phonetics from the domain of langue are facets of Saussurian theory that have been at least in practice correctly rejected in much of modern structuralism. The conclusion from this, namely that it is unmotivated to regard linguistic innovations as aspects of parole, was duly drawn by Jakobson (1929) who emphasized the systematic aspect of linguistic change and observed that the ultimate historical source - as distinct from theoretical basis - of Saussurian diachrony is to be found in the neo-grammatician doctrines of the 19th century.

The same historical dependence on the neo-grammarians is even more striking in the case of American structuralism (cf. Postal, forthcoming). Bloomfield in the main accepted Paul's position (explicitly rejecting, however, his reliance on facility of articulation, and making no mention of imperfect learning). But his way of justifying it is quite different from that of Paul and Saussure. The first argument which Bloomfield offers in favor of the neo-grammatician theory is that it led to discoveries like Grassmann's Law (1933, 355) and Verner's Law (359). Alchemy, too, led to many important discoveries; and besides, Bloomfield's argument is historically false as neither Grassmann nor Verner are known to have held the neo-grammatician position either implicitly or explicitly.

The second, much more interesting argument is the principle of separation of levels:

"Theoretically, we can understand the regular change of phonemes, if we suppose that language consists of two layers of habit. One layer is phonemic: the speakers have certain habits of voicing, tongue-movement, and so on. These habits make up the phonetic system of the language. The other layer consists of formal-semantic habits: the speakers habitually utter certain

combinations of phonemes in response to certain types of stimuli, and respond appropriately when they hear these same combinations. These habits make up the grammar and lexicon of the language." (364-5)

If this statement is to make any sense as a defense of the neo-grammarians position, it must be interpreted to mean not merely that there are two layers, but also that the phonemic layer is independent of the grammatical-lexical layer in the sense that the phonological structure of utterances in languages is not determined in part by their grammatical-lexical structure. But that is an untenable view, as Bloomfield in effect acknowledged on the very next page, where he observed that some speakers of American English use "a shorter variant of the phoneme [a] ... before the clusters [rd, rt] followed by a primary suffix [-r, -n], as in barter, Carter, garden, marten (Martin). Before a secondary suffix [-r, -n], however, the longer variant is used, as in starter, carter ('one who carts'), harden; here the existence of the simple words (start, cart, hard), whose [a] is not subject to shortening, has led to the favoring of the normal, longer variant. The word larder (not part of the colloquial vocabulary) could be read with the shorter variant, but the agent-noun larder ('one who lards') could be formed only with the longer type of the [a]-phoneme." Bloomfield explicitly pointed out that long and short [a] are "sub-phonemic variants", being predictable from the nature of the suffix and hence in spite of such minimal pairs, not distinctive.<sup>4</sup> But diachronically, Bloomfield did not come to the obvious corresponding conclusion that the shortening failed to take place before the secondary suffixes but rather assumed that it took place even there, with the long allophone later being introduced analogically on the basis of the corresponding simplexes.

Thus Bloomfield still took the principle of separation of levels much more seriously in diachronic phonology than in synchronic phonology, ignoring it in this and many other instances in his actual descriptions while simultaneously insisting on it in an exactly parallel historical situation. The contribution of his disciples was to take seriously the principle in synchrony as well.<sup>5</sup> Thus the neo-grammarians theory of sound change is indubitably one source of taxonomic phonemics in America.

At any rate, in spite of various uncertainties of interpretation, we can say that the neo-grammarians, Saussurian, and Bloomfieldian theories of sound change place equivalent or very similar limitations on what kinds of sound change may happen. The theoretical justification for this shared position - let me call it the taxonomic position - is in each case based on assumptions which for independent reasons are untenable: in Paul, the absence of any level of higher-level, non-phonetic representation; in Saussure, a particular view of langue and

parole; in Bloomfield, a form of separation of levels. Thus, there is no reason to believe that these limitations on phonological change are correct.

But besides these theoretical arguments we can bring to bear evidence of an empirical kind on the question. The taxonomic position, in attributing phonological change to non-linguistic factors, must exclude two kinds of cases in particular: phonetic leaps other than the special cases of metathesis, assimilation, dissimilation, etc., and phonological changes conditioned by non-phonetic grammatical environments. To exclude such cases in generative grammatical theory, on the other hand, would be possible only on an ad hoc basis. Thus, if such changes do exist, that fact would strongly favor generative grammatical theory over taxonomic theories of linguistic change. In the following I would like to add to the presently available evidence that shows that phonetic leaps and non-phonetically conditioned sound changes indeed occur.

#### 1.4 Phonetic leaps

A common type of case where phonetic leap seems indicated is the interchange of velars and labials found in many languages (Jakobson, 1938). For example, in Rumanian, velar stops (including nasals) have become labial before dentals by the rule

$$(16) [-\text{continuant}] \rightarrow [-\text{compact}] / \quad \begin{bmatrix} [-\text{continuant}] \\ [-\text{grave}] \end{bmatrix}$$

Cf. Latin lignum, pūgnūm (with phonetic ŋn); Rumanian lemn, punn; Latin frictum, lactem, octo; Rumanian fript, lapt, opt. Many other cases might be cited, e.g. the Greek change of y to t after labials (cf. kʰalepto from kʰalepyō). But in such examples it is always possible to dream up some roundabout series of intermediate steps which make the change compatible with articulatory drift or replacement.<sup>6</sup> Cases where this is ruled out are thus necessarily of a rather special kind.

A surprising type of phonetic leap where taxonomic theories would seem to run into insuperable difficulties is pointed out by Halle (forthcoming). To make clear what kind of process it involves, a brief digression on a type of phonological rule is necessary. The usefulness of variable feature specifications is not limited to assimilations such as rule (7), dissimilations, feature agreement, etc. There is also synchronic motivation for rules of the type

$$(17) [\alpha F] \rightarrow [-\alpha F]$$

which exchange the marking of a feature in some segment,<sup>7</sup> rather like castling in chess. For example, as a result of Verner's Law, voiceless continuants alternated in Old High German with the corresponding voiced continuants, phonetically realized as stops by the rule

$$(18) \begin{bmatrix} + \text{voiced} \\ + \text{obstruent} \end{bmatrix} \rightarrow [-\text{continuant}]$$

Cf., with x:g, zichan 'draw', gizogan 'drawn'; with s:r, kiosan 'choose', gikoran 'chosen'. But in the same environment there was a corresponding d:t alternation, e.g. siodan 'boil', gisotan 'boiled'. No separate rule corresponding to Verner's Law for dentals was needed as d and t were derived from θ and ð, respectively, by cyclical application of (18) and an additional rule (19) which exchanged the voicing of nonstrident dentals:

$$(19) \begin{bmatrix} + \text{obstruent} \\ - \text{grave} \\ - \text{strident} \\ \alpha \text{ voiced} \end{bmatrix} \rightarrow [-\alpha \text{ voiced}]$$

Thus:

sioθan	gisodan	
"	gisodan	by (18)
siodan	gisotan	by (19)
siodan	"	by (18)

The existence of exchange rules like (19) has an interesting diachronic consequence. In generative grammatical theory there is no natural way of excluding such exchange rules from being adopted as innovations. But within taxonomic theory the possibility of a sound change which reverses the marking of a feature is necessarily excluded no matter which form of change we take. Hence the question whether exchange rules exist as innovations bears directly on the larger issue of how to view linguistic change in general. Halle has suggested that such an innovation may in fact have been part of the Great Vowel Shift. I would like to cite two further examples in each of which the notion of an exchange rule appears to offer the key to a traditional problem of historical linguistics.

The ten-vowel system of Classical Latin was reduced in Vulgar Latin as indicated in the following table of correspondences.

(20) Cl. L.	I	i	ē	e	ā	a	o	ō	u	ū
V. L.	i	e	ɛ	ə	a	ɔ	o	ʊ	u	u

The most straightforward assumption is that this development resulted from two changes. The first was a lowering of the short vowels e, o to ɛ, ɔ and i, u to e, o, respectively:

$$(21) \begin{array}{l} [-\text{diffuse}] \rightarrow [+ \text{compact}] \\ [-\text{diffuse}] \rightarrow [-\text{diffuse}] \end{array} / \quad \boxed{\begin{array}{l} -\text{consonantal} \\ -\text{long} \end{array}}$$

The second was the loss of the original quantitative distinctions in vowels, by some rule which made quantity contingent on syllabic structure in the well-known manner, perhaps

$$(22) [-\text{consonantal}] \rightarrow \begin{cases} [+ \text{long}] \\ [- \text{long}] \end{cases} / \quad \boxed{\text{CC}}$$

Thus I follow Spence (1965) in rejecting the 9-vowel system which is usually (quite gratuitously, so far as I can see) reconstructed as an intermediary stage between Classical and Vulgar Latin. It will not matter, however, for the purposes of the following discussion whether this is correct; nor will it matter whether (21) and (22) led to restructuring and the second row of (20) represents the phonemic vowels of Vulgar Latin, as is traditionally assumed, or whether, as is convincingly argued by Schane (1965), Foley (1965) the underlying vowels of Latin remained with counterparts of (21) and (22) still functioning as morpho-phonemic rules even in the modern Romance languages.

In the dialects of Northern Corsica<sup>8</sup>, the vowels of Vulgar Latin (preserved essentially unchanged in Italian) developed further in the following way:

$$(23) \begin{array}{ccccccc} i & e & \cancel{\varepsilon} & a & \cancel{o} & \cancel{\circ} & u \\ | & & \times & | & & \times & | \\ i & e & \varepsilon & a & o & \circ & u \end{array}$$

That is, mid and open vowels have exchanged places with each other. Cf. the following examples:

(24)	Classical Latin	Standard Italian	N. Corsican
	dictus	detto	dətto
	pectus	pətto	petto
	musca	mosca	məsca
	noctem	nɔtte	notte

Evidently N. Corsican has undergone the sound change

(25)	<table border="1"> <tr><td>- consonantal</td></tr> <tr><td>- diffuse</td></tr> <tr><td>α round</td></tr> <tr><td>α grave</td></tr> <tr><td>β compact</td></tr> </table>	- consonantal	- diffuse	α round	α grave	β compact	→ [-β compact]
- consonantal							
- diffuse							
α round							
α grave							
β compact							

This is the case where taxonomic theories fail outright. We can neither attribute the switch of mid and open vowels to imperfect learning nor to articulatory drift, as this would leave it a mystery why the original distinction between mid and open vowels has nowhere been lost in the switch. No-one would wish to say that some diacritic feature - say nasalization - was introduced to distinguish mid and low vowels before their interchange and later disappeared, leaving no trace behind. Nor would it make matters any easier to derive the N. Corsican vocalism directly from Classical Latin without postulating an intervening Vulgar Latin ancestor for this dialect. On the contrary this would merely complicate the account in other ways without contributing in any way to the solution of the fundamental problem posed by this example for taxonomic theories. We should still have to explain how Classical Latin i became N. Corsican ɛ (dictus > detto) without merging with e (pectus > petto) "on the way". But now there would be the additional implausible consequence of having to say that Classical Latin i and ɛ came to fall together in N. Corsican in some quite different way than elsewhere in Italian and in Italo-Western Rómance in general.

However, not only does the present example provide unanswerable counter-evidence to existing taxonomic theories of sound change. It is also true - and here lies the deeper significance of this and other similar examples - that any theory of sound change based on a taxonomic conception of linguistic structure must deny the possibility of a switch of this kind. That is, no such theory can even reach the level of descriptive adequacy as it cannot accommodate such empirical facts as this example. The only basis on which they become even conceivable is provided by the concept of an abstract underlying representation to the phonetic representation by rules of a certain abstract kind. But furthermore, as we have seen, generative grammar not only provides a general basis for describing switching sound changes like this but also explains their existence in the following specific sense. The fact that such innovations should exist is seen to be a joint consequence of two independently motivated assumptions about language: first, that phonological rules with variably specified features are part of language, and second, that grammars change by receiving new rules. That these two assumptions should lead to such a surprising, previously quite unsuspected but correct consequence lends further strong support to both.

The following example is in many ways similar. Vogt (1958) divided the dialects of Armenian into six groups according to their treatment of the stops. A seventh group - the most interesting of all - was for some reason omitted in Vogt's study but included in subsequent papers by Garibian (1959) and Djaukian (1960). The data are summarized in the following table, in which the dialect groups are named roughly to indicate their location and the stops are represented by the dental series:

Vogt	Garibian	Indo-European	t	d	dh
			th	t	d
I	II	East Central	th	t	dh
II	I	West Central	th	d	dh
III	VI	Northern	th	t	d
IV	VII	Eastern	th	t	t
Va	III	Western	th	d	d
Vo	V	Northwestern	th	d	th
--	IV	Southern	th	d	t

With a single exception, it is fairly clear how these dialect differences arose. In the Central dialects (both East and West Central) the Old Armenian voiced stops were aspirated (e.g. d > dh) by the rule

$$(26) \quad [+ \text{voiced}] \rightarrow [+ \text{tense}]$$

The resulting voiced aspirates were then devoiced to yield the Northwestern type of dialect with th, which also is found in scattered enclaves in the West Central area. Two further phonetic changes then spread across the area, one coming from the West and the other coming from the East. The Western and West Central dialect underwent a voicing of the unaspirated stops (t > d) by the rule

$$(27) \quad [- \text{tense}] \rightarrow [+ \text{voiced}]$$

The West Central dialect thus underwent both (26) and (27), in that order. The Eastern dialect underwent just the opposite change in which unaspirated stops were devoiced (d > t) by the rule

$$(28) \quad [- \text{tense}] \rightarrow [- \text{voiced}]$$

This leaves the Southern dialect, in which d and t have exchanged places. It is suggestive that this dialect borders both on the Western area where (27) turned t to d and on the Eastern area where (28) turned d to t. It seems reasonable to ascribe the Southern switch of t and d to combined participation of this group in both these sound changes, and this is in fact the solution proposed by Djaukian (1960, 48). But

it will not do to suppose that the two rules operated in sequence since that would have led to a merger of t and d no matter which order of application we assume.

A solution to this impasse is the following. Suppose that the Southern dialects were reached simultaneously by (27), spreading from the West and (28) spreading from the East. The two rules were then combined into a single rule

$$(29) \quad \begin{bmatrix} [-\text{tense}] \\ [\alpha \text{ voiced}] \end{bmatrix} \rightarrow [-\alpha \text{ voiced}]$$

which accomplished the switch in a single step. In this way rules with variably specified features not only provide a rational account of this seemingly impossible phonological change, but also plausibly relate it to the changes in the neighboring dialects.

It must be pointed out, however, that unlike the Italian case there are here other accounts of the change which do not involve a rule with variably specified features, although none of the alternatives are compatible with taxonomic notions of sound change. There is one in particular which becomes plausible if we accept the proposal, made independently by Vogt (1958) and Benveniste (1959) that the Old Armenian graphic voiced stops d etc. were actually aspirated dh etc. If Vogt and Benveniste are right, the above picture of the historical development of the Armenian dialects must be revised. The central dialects would then turn out to be conservative rather than innovating in alone retaining original dh which is deaspirated elsewhere. The Southern dialect would then be the result of two innovations: first a change of t to d by rule (27), shared with the Western dialects as before, and a second, two-feature change of dh to t by a rule

$$(30) \quad \begin{bmatrix} [+ \text{tense}] \\ [+ \text{voiced}] \end{bmatrix} \rightarrow \begin{bmatrix} [- \text{tense}] \\ [- \text{voiced}] \end{bmatrix}$$

shared with the Eastern dialect.

In this interpretation of the Old Armenian consonant system the development of the Southern dialect is just as clear a case of sound change by phonetic leap. It is necessary to suppose that in the Southern dialect dh became t by a change of tenseness and voicing at once since otherwise it would have merged with either d or th "on the way". In fact, a two-feature change of some kind or a switch is required to get from IE t d dh to Southern Armenian th d t no matter what intermediate Old Armenian stop system we reconstruct, unless we wish to introduce a deus ex machina in the form of some new feature that makes its appearance before the change and later is removed.

An incidental feature of the present example is that it highlights the pointlessness of a structural dialectology that draws so-called typological rather than historical isoglosses, that is, distinguishes dialects according to points of structural difference rather than according to the innovations through which they diverged. The futility of the former approach is a consequence of the fact that it is innovations and not the resulting structural changes that spread and in which (with the exception of the cases to be discussed in Chapter 2) the significant regularities about linguistic change are to be found. It is entirely possible both that the same innovation has different consequences in different grammars and that different innovations have the same consequences in different grammars. If in the present example we were to divide the dialects into those with two stop series and those with three, we would be linking together dialects that have nothing to do with each other and separating dialects that are closely related. We would have to group together the disjoint Eastern and Western dialect to the exclusion of the intervening Central dialects because Old Armenian t and d have merged in the former two even though the merger has come about in an entirely different way in each. On the other hand, we could show no relation at all between the contiguous Western and West Central dialects although both have undergone (27). They would have to be separated because the rule had a different effect in the two grammars, owing to interaction with other rules. Instead of a few well-defined isoglosses we would this way end up with either just two completely discontinuous dialect groups with widely interspersed members if we took the number of stop series as an isogloss, or else with an unmanageable welter of dialects if some more detailed point of structural difference is chosen instead.

### 1.5 Non-phonetic environments

Thus one consequence of the taxonomic view of linguistic change, that sound change is restricted to articulatory drift and replacement is seen to be false because of phonetic leaps. Another respect in which it is challenged by generative grammatical theory is in its denial that non-phonetic grammatical factors - elements of langue - can play a role in sound change. The fact that the denial of non-phonetically conditioned sound change follows from viewing change as part of parole was seen clearly by Saussure:

"A phonetic change is also unlimited and incalculable in that it affects all types of signs, making no distinction between radicals, suffixes, etc. This must be true a priori, for if grammar interfered, the phonetic phenomenon would mingle with the synchronic fact, a thing that is radically impossible. It is in this sense that we can speak of the blind nature of the evolutions of sounds." (152)

There are at least three kinds of non-phonetic factors that can condition sound change:

1. boundary markers
2. constituent structure
3. underlying phonological representations

It is extremely frequent that syntactic boundaries (word and morpheme boundaries) play a role as conditioning factors of sound change. Denial of this has been possible because very often alternative descriptions within the taxonomic framework can be supplied if the device of analogy is introduced in ad hoc ways. This is true particularly in two types of cases.

First, particularly when the role played by the boundary is to block a change, it is often possible to say instead that the sound change operated regardless of boundaries and to suppose that forms where the change did not operate result from analogical restoration under the influence of other forms. As we have seen, this is how Bloomfield explained why length is retained in nouns with secondary suffixes like carter 'one who carts'.

Secondly, in other cases it is possible to say that the real environment of the sound change is not the abstract boundary but some phonetic property that is associated with this abstract boundary. For example, if in a language obstruents become voiceless before a word boundary the fact that word boundaries are optionally realized as phonetic pauses allow one to salvage the taxonomic theory by saying that devoicing takes place before phonetic pause and that the forms thus devoiced become generalized to cases where the phonetic pause is not present.

The only examples that can carry conviction under these circumstances are those where it is not possible (by anyone's standards) to invoke analogy and where, at the same time, the boundary cannot be assumed to have a phonetic reflex such as pause. Such a case is found e.g. in Greek, where velars are voiced before a morpheme boundary followed by a nasal (Lejeune 1947, 67). Cf., with voicing, Perf. Mid. Pass. forms like pépleg-mai, peplég-metha, pepleg-ménos (plék-ō), derivatives like plég-ma, eug-ma (éukhomai), deig-ma (déiknumi), árg-ma (árkhō); but with retention of voicelessness, kékmēka (kámnō), akmē, ákmōn, dokhmós.

A second major aspect of grammatical structure available to phonological rules is categorial information, that is, labeling of constituents as nouns, verbs and so forth. In synchronic phonology, categorial information plays a central role, particularly in accentual and inflectional systems, and this fact alone would make it very surprising

if phonological change was always immune to it. There is not much point in arguing this at length as the literature teems with unchallenged examples which appear to me to establish beyond any doubt the fact that categorial information can condition sound change (Allen 1953, Lindgren 1953, Manczak 1960, Cohen 1939, Birkeland 1952). A typical case (typical also for its lack of interest) is the following. In Estonian, final -n has been lost except as the first person singular morpheme. (Kettunen, 1929). Thus Balto-Finnic and modern Finnish kanōan 'base (Gen.)' becomes Estonian kanna while Balto-Finnic and modern Finnish kanōan 'I carry' has retained its final -n in Estonian kannan. Note that there can be absolutely no question of analogy here as the environments in which first person -n is found are in no way distinct from the environments in which other final -n's are found.

Of greater interest is the fact that not only surface syntactic structure but also underlying phonological representations may function as conditioning environments of sound change. Clear instances of this are much harder to find as they usually involve the need for determining relative chronology of rules in a conclusive way, a thing which is generally very difficult to carry out. In the following example this can be done with a fair degree of finality.

In Latin, vowels become lengthened before clusters of the form "voiced obstruent + voiceless obstruent" by Lachmann's Law (31). On the phonetic level, however, such clusters are eliminated by a subsequent rule of regressive assimilation of voicing (32).

- $$(31) [-\text{consonantal}] \rightarrow [+ \text{long}] / \quad \begin{array}{c} [+\text{obstruent}] \\ [+\text{voiced}] \end{array} \quad \begin{array}{c} [+ \text{obstr}] \\ [- \text{voiced}] \end{array}$$
- $$(32) [+ \text{obstruent}] \rightarrow [\alpha \text{ voiced}] / \quad \begin{array}{c} [+\text{obstruent}] \\ \alpha \text{ voiced} \end{array}$$

Thus underlying ag-tus becomes first āgtus by (31) and then āktus by (32), while an underlying fak-tus is not affected by either rule.

From the point of view of taxonomic theories of sound change, the example is embarrassing because (31) is a Latin innovation and (32) is a Proto-IE heritage in Latin. Lachmann himself was not aware of any problem here<sup>10</sup>, but to the neo-grammarians the difficulty had become clear:

"Da schon im Indogermanischen ein stimmhafter Konsonant vor einem stimmlosen den Stimmtón verlieren muss, bereitet die Verschiedenheit in der lateinischen Entwicklung von idg. \*aktos und idg. \*d̥lēktos Schwierigkeiten, die auf irgend eine Weise beseitigt werden müssen." (Streitberg 1915, 195).

The validity of this example depends crucially on the correctness of the given relative chronology of (31) and (32) in Latin. One might try to dispose of the example in two ways: by pushing Lachmann's Law in some form back into Indo-European, or by claiming that the voicing assimilation took place not in IE but only within the separate development of Latin itself.

Assimilation of voicing in obstruent clusters is so common a phenomenon that its ubiquity in the IE languages cannot alone guarantee its antiquity. It is a priori conceivable (though certainly very unlikely) that (32) was not proto-IE but was later independently adopted by each separate IE language, and that in the case of Latin this took place after (31) was adopted.

Saussure (1889) saw that this possibility is out of the question because certain other rules whose proto-IE status is clear presuppose the prior application of (32). One is the simplification of geminate consonants before consonants as in Skt. dus-stutíś > dustutíś, Latin dis-stō > disto, Gothic twis-standan > twistandán. This in turn preceded the assimilation of dentals before dentals as in dad-dhi > dad<sup>z</sup>dhi (cf. Avestan dazdi). The relative chronology is nailed down by the following kinds of derivations which must be reconstructed for proto-IE:

(33)	underl. form	sed-tlom	'seat'	sed-tos	'sit (p.p.)'
(32)		settлом		settос	
geminate simpl.		setлом		-	
dental assibil.		-		set <sup>s</sup> tos	

Cf. Skt. satrám, Gmc. \*se<sup>h</sup>la (OHG sēdal) 'seat'; Lat. ob-sessus, Av. hastō 'sit (p.p.)'. For the rules, cf. Meillet (1937, 130-1).

One might also think of denying that (31) is a Latin innovation and push it back in one form or another into IE, S.-Y. Kuroda has suggested the following ingenious attempt along these lines (though this must not be taken to mean that Kuroda is necessarily a neo-grammarian). Suppose that IE had a modified form of (31) which made IE vowels half-long before clusters of the form "voiced obstruent + voiceless obstruent". Phonetically this is certainly plausible in view of the similar lengthening of vowels before voiced consonants in English and many other languages. Subsequently - so Kuroda proposes - (32) was added with the result that the half-long vowels merged with the short vowels in all IE dialects except Pre-Latin, where they merged with the long vowels.

But this proposal fails because the lengthening does not take place in all words which had the requisite clusters in IE but just in

those which had them in Latin. There is no lengthening, for example, in tussis 'cough', lassus 'tired', lectus 'bed' although these words had IE roots with voiced obstruents \*tud (Skt. tudati), \*ləd (Greek lēdein), \*legh (Greek lekhos). The reason the vowel stayed short in these Latin forms is that the root had either disappeared in Latin (as in the case of the latter two roots) or else been dissociated from its derivative (as in the case of tud which survives in tundo). Hence there was no morphophonemic retention of voicing in the underlying representations of tussis, lassus, lectus and consequently no lengthening by Lachmann's Law. Lachmann's Law thus historically postdates these lexical changes in Latin and hence is not IE but a relatively late Latin innovation.

There is yet another way of accounting for Lachmann's Law within the taxonomic framework, which is due to Saussure (1889; 1916, p. 168). By postulating an analogical restoration of voicing in actus Saussure tried to explain the lengthening on the basis of a purely phonetic environment without denying that (31) is a relatively Late Latin innovation and (32) is proto-IE. He supposes that IE aktos reverted to phonetic agtos, then was lengthened to āgtos and finally reassimilated to aktos. In spite of its ad hoc character and phonetic implausibility (on which Saussure himself remarks) this has come to be the generally accepted view (Sommer 1948, 122; Streitberg 1915; Hirt 1921, 308; Niedermann 1953, 78; Meillet-Vendryes 1963, 32). But there are insurmountable objections to it. To account for lengthening in dental stems (e.g. cāsus) we should then also have to suppose that forms like "cad tus" were restored, and that after the lengthening by Lachmann's Law these forms underwent not only reassimilation of voicing but also reassibilation by the old rule that dentals became sibilants before dentals. This kind of miraculous repetition of history stretches our credulity to the breaking point. It snaps when we recall that dental clusters of secondary Latin origin do not in fact assimilate in Latin, e.g. ad-terō > atterō, and not "asserō". There is, so far as I can see, no way of saving Saussure's theory of Lachmann's Law.

It is not possible to say, of course, that there might not be still another account compatible with taxonomic theory that we have not countered here, although it is very likely that the above exhausts all the reasonable ones. This example well illustrates how rich an array of facts and theoretical considerations may have to participate in an act of linguistic reconstruction if it is to have historical validity. In order to determine the prehistory of Latin with regard to just the two simple rules (31) and (32) it turned out to be necessary to relate them to many other rules both in Latin and in other IE languages, to consider the treatment of isolated forms both as negative and positive evidence (again, in Latin as well as in other IE languages), to take account of the possibility of analogy and of borrowing of rules. It is really pure luck when in a case such as this the requisite crucial data is available.

### 1.6 The placement of innovations

From the fact that new phonological rules can be placed elsewhere than at the end of the phonological component, as shown by the examples of section 1.5, it certainly does not follow that they can be placed everywhere. From the inception of historical studies in the present framework it has been clear that there are heavy restrictions on the placement of innovations. Halle (1962) proposed the condition that

"linguistic change is normally subject to the constraint that it must not result in the destruction of mutual intelligibility between the innovators - i.e., the carriers of the change - and the rest of the speech community."

As Halle immediately notes, a very clear type of innovation which disturb intelligibility are secret languages such as Pig Latin. Thus it is necessary to say that the intelligibility condition does not delimit innovations which are inherently possible for mature speakers, but rather innovations which have a chance of survival in a speech community. The constraints on linguistic change are thus regarded as evolutionary in the sense that they impose no limit on the way a speaker can modify his grammar but rather prevent unfit innovations (i.e. those that prevent communication between innovating and conservative dialects) from gaining acceptance and becoming permanent changes.

One thing that this fails to account for is the fact that secret languages are usually derived by rules applying to a late level of representation. It would be easy to do it by adding appropriate rules high up in the phonology, but this does not seem to occur in actual cases. For example, a rule making all vowels tense would wreak havoc in English morphophonemics if applied before the Great Vowel Shift rules. (Applied on the phonetic level it would convey the effect of a French accent but would have no serious effect on intelligibility). Such an innovation is far harder to acquire than the Pig Latin rule, and it is not unreasonable to suppose that this is due to some inherent limitation on human capacity for modifying language.

On the other hand, it seems that innovations which destroy or at least impair intelligibility are by no means necessarily doomed to be abortive. Surely the interchange of nondiffuse vowels or the interchange of voiced and voiceless consonants is no less disruptive to comprehension than secret language rules. Other examples are perhaps the explosive breakdown of Old Irish phonology in the 7th and early 8th centuries (Watkins, forthcoming), and some stages of the Great Vowel Shift. Moreover, it is difficult to understand the origin of phenomena such as infixation which are quite like secret language rules in character if it is assumed that such rules cannot survive.

Finally, the intelligibility condition fails to explain the fact, observed by Halle (1962), that new rules tend to get placed "at points where there are natural breaks in the grammar, e.g., before the first morphophonemic rule involving immediate constituent structure of the utterance (i.e., at the end of the morpheme structure (MS) rules); at the end of the cyclical rules which eliminate the immediate constituent structure of the utterance from the representations; before the phonological rules that eliminate boundary markers (junctions) from the representation", at the end of the phonological component, and presumably at the point before the binary features assume n-ary values. Although there are exceptions to this, as Halle points out, the rarity of cases where a rule is stuck into the middle of a sequence of ordered rules is, along with their indubitable existence, a fact that must be accounted for in a theory of linguistic change. I do not see how this is possible with the intelligibility condition, as it is not necessarily true that breaking a sequence of rules in this way must impair intelligibility in any significant way. For example, we would surely want to exclude the possibility of a rule making vowels tense before t being entered before the Great Vowel Shift rules and the consonantal morphophonemics, in spite of the relatively minor effect such an innovation would have on intelligibility.

It appears, then, that there are both possible innovations that disrupt intelligibility and impossible innovations that would maintain it. This suggests that the true constraints may be found elsewhere, perhaps in some inherent psychological limitations of mature speakers to modify their grammars. Such a condition should evidently be formulated in terms of the structure of the grammar itself. Along these lines, I would like to offer the following speculations.

Rules can be thought of as consisting of two parts, the structural analysis and the structural change. Thus, if we write the Indo-European voicing assimilation rule (32) in the form

$$(34) \quad [+ \text{obstruent}] \quad \begin{array}{c} [+ \text{obstruent}] \\ \downarrow \\ [- \text{voiced}] \end{array}$$

[ $\alpha$  voiced]

the top line constitutes the structural analysis and the bottom line the structural change (Cf. McCawley, 1966). In the more usual notation for phonological rules (Halle, 1962), which I am using here, the structural change is the part between the arrow and the slash, and the structural analysis is the remainder of the rule.

Looking at innovations in terms of this distinction, we see that it is apparently possible to place a rule with the structural

analysis A at a point before some rule removes representations satisfying A (this is the case in Lachmann's Law) or introduces representations satisfying A (cf. several examples in Postal, forthcoming). We can now ask whether it is conversely possible to place a rule introducing or removing representations that satisfy A before a rule with the structural analysis A. I know of no clear examples of this latter kind. Take the schematic phonological rules (35i) and (35ii) where  $\alpha, \beta$  are + or -, A is a feature, and the empty brackets can be assumed to contain sets of specified features.

- (35) .(i)  $[\alpha:A] \rightarrow [\quad]$
- (ii)  $[\quad] \rightarrow [\beta A]$

Into a grammar with (35ii) it is apparently possible to place an innovation (35i) to precede (35ii). But it is a completely open question whether a grammar with (35i) could receive before that rule the rule (35ii).<sup>12</sup>

It is psychologically very plausible that the latter type of innovation would in fact be impossible. It would then be true that innovations could not disrupt the derivations of a language by affecting the application of other rules to any representations of the language. In effect, it would be the case that underlying representations could condition sound change but not be changed themselves in a way that would affect the operation of later rules. The question is worth pursuing, I think, both for its psychological interest and for its immediate relevance to the determination of relative chronology in linguistic reconstruction.

## Chapter 27

Imperfect Learning2.1 The Problem

If innovation and restructuring as described in Chapter 1 were the only form of linguistic change, we should expect isoglosses (distinguishing features of dialects) to consist typically of the absence versus presence (or, if restructuring has taken place, former presence) of rules. In view of the evidently very severe restrictions on the placement of innovations (whatever the exact nature of these restrictions may be - cf. section 1.6), which suggest that the placement of an innovation at different points of a grammar should be a rare form of dialectal differentiation, we should not expect differences in the order of rules to figure as isoglosses very frequently. As any description of a group of closely related languages or dialects will show, these expectations are not confirmed by the facts. In practice, isoglosses very often consist of small differences in the detail of essentially shared rules, or differences in the order of application of shared rules. For example, M. Halle and T. Lightner (forthcoming) show the striking fact that all Slavic languages share basically the same rules relating to the several palatalizations of velars before front vowels, but that various groups of Slavic languages differ in the order in which these rules must apply. G. H. Matthews (forthcoming) shows that many of the modern Siouan languages have a set of rules relating to the obstruent system which vary in relatively minor ways from language to language, and has proposed an explanation for the origin of differences, to which I return below. W. O'Neil, in some unpublished work, has demonstrated that similar differences, both with regard to order and form of rules, underlie the dialectal diversity of Old English. Halle (forthcoming a) has described the rule that determines how unstressed vowels coalesce in Russian, and shown variation of a relatively minor kind in this rule from dialect to dialect.

There is always a possible explanation for such differences within the theory of Chapter 1 as long as we look at dialects without knowledge of their historical origin. The explanation for order differences is that rules have spread from different directions and crossed. For example, if rule A spreads from West to East, and rule B spreads from East to West across some dialect, then, if the two rules are critically ordered with respect to each other, the Western area will end up with the order (A, B) and the Eastern area with the order (B, A). Such geographical crossing of rules is in fact very common. Small differences in the form of rules could conceivably be due to changes as rules are borrowed from one dialect into another. However, we find

just the same types of difference when comparing successive stages of the same dialect rather than geographically adjacent dialects, and here the wave theoretical and "imperfect borrowing" explanations are excluded.

In this chapter I will try to justify these remarks and propose a form of linguistic change that contributes to the formation of such dialectal differences. The chapter is organized as follows. In this section I cite some examples and bring up some of the difficulties that arise when the notion of innovation is extended to cover them. The next two sections (2.2-3) suggest an alternative solution, which then in 2.4 is compared with more traditional taxonomic accounts. A final section picks up some remaining loose ends.

Ex. 1. In many Swiss German dialects (Jutz, 1931, 116) there are alternations of the following kind between ä and ǣ.

- (1) ä 'egg' ~ ǣiər 'Pl.'  
tswā 'two' ~ tswǣiər 'two-deciliter bottle'  
srā 'scream (n.)' ~ srāiə 'scream (v.)'

We derive ä from the underlying diphthong by the Monophthongization rule

- (2) ǣ → ä except before vowels.

Rule (2) is critically ordered with respect to the umlaut rule, by which vowels become nongrave (fronted) in certain environments such as Plurals, Diminutives, etc.

- (3) [- cons] → [- grv] / ...

We know that (3) is a much older rule, historically, than (2) and that synchronically, too, (3) at one time preceded (2) in the grammars of these dialects.<sup>1</sup> Thus ä from ǣ remained unchanged in umlauting environments, e.g. äli 'egg (Dimin.)', srā 'scream (Pl.)'.

But in most dialects<sup>2</sup> the original situation has been modified in that ä from ǣ does undergo umlaut to ǣ. In these dialects, äli, for example, has changed to ǣli. In terms of the grammars this means that the order

1. Umlaut (3)
2. Monophthongization (2)

<sup>1,2</sup> has been replaced by the order

1. Monophthongization (2)
2. Umlaut (3)

It is by no means impossible, technically, to account for Ex. 1 within the framework of Chapter 1. There are even two ways of doing it. One is to assume that a rule (2'), identical with (2), was entered before (3). This would have the effect of making (2) vacuous, causing it to be dropped when the resulting grammar is subjected to the evaluation measure. The other is to assume that a rule (3'), identical with (3), was entered after (2). The optimal grammar for the resulting output would again be the same as above.

The unfortunate aspect of this is the arbitrariness of the choice between the two descriptions. It is hard to see how the distinction between them could correspond to any linguistic difference. The two distinct grammars containing an identical rule at two different points which are required as "virtual" intermediaries seem to be mere artifacts made necessary by an inappropriate extension of the notion of innovation to account for a quite different kind of process.

The difficulties become considerable in an example such as the following (Ex. 2). The rule that word-final obstruents are devoiced, once common to all dialects of German, has been lost in some dialects of Northern Switzerland<sup>3</sup> as well as in some varieties of Yiddish<sup>4</sup>. In place of bunt:bundes they have bund:bundes with the morphophonemic distinction of voicing now appearing phonetically in word-final position. We know that they indeed once possessed the devoicing rule, as it has permanently affected a handful of isolated words like (a)vek 'away'. These had, historically, a voiced final obstruent but lost it even morphophonemically after the phonetic devoicing took place as retention was not motivated by any related forms. Hence there was also no basis for reintroducing the voicing in these words. It is possible, with some ingenuity, to treat even this case as the result of an added rule. We cannot, clearly, simply suppose that a late rule making final obstruents voiced was added. Such a rule could not distinguish between morphophonemically voiced and voiceless stops and would wrongly turn into bund not only the bunt that is related to bundes but also the bunt that is related to buntes. But if morphophonemically voiced obstruents were somehow provided with a diacritic feature before they got devoiced, this diacritic feature could later serve as an environment for revoicing. For example, before the devoicing final voiced obstruents could be made, say, glottalized, and after the devoicing glottalized consonants could be made non-glottalized and voiced. Obviously this bears not the faintest resemblance to what actually happened and no-one would want to salvage a theory of linguistic change at the price of such an absurd description.

Another proposal that has been considered<sup>5</sup> is that such changes instead arise as instructions of speakers not to apply certain rules. This already constitutes a departure from the decision to restrict linguistic change to added rules. But it is not clear that it is a departure in the right direction. It asserts, in effect, that some linguistic changes (such as Ex. 1) arise in langue, while others (such as Ex. 2) are matters of parole at first, and become part of langue only as the language is transmitted to a succeeding generation of speakers. There seems to be no good reason for departing so radically from the traditional view and making such a deep distinction between what would appear to be phenomena of a rather similar kind.

Now consider Ex. 3, which again concerns the Umlaut rule. As noted in 1.2, this rule in its original form turned vowels nongrave and in addition made short vowels non-compact. But in many dialects of German, the productive umlaut of a is ä instead of e as originally. For example, in the Low German dialect of Prignitz (Mackel 1905-7), we have

(4)	gast	gäst
	kraft	kräftig

and so forth, instead of the expected

(5)	gast	gest
	kraft	kreftig

However, e which is not synchronically derived from a has remained unchanged. This includes not only original Germanic e in words like nest but also e from umlauted a in words like bet 'bed', net 'net' where e has become phonemic since there is no reason to derive it synchronically from an underlying a. These dialects have changed in that the umlaut rule has become simplified from its original form

$$(6) \quad \left[ \begin{array}{l} \text{- consonantal} \\ \text{-- long} \end{array} \right] \rightarrow \left[ \begin{array}{l} \text{- grave} \\ \text{-- compact} \end{array} \right] / \dots$$

to the form (3).

In this case there is, as in the previous example, no way of characterizing this change by addition of rules, short of, again, some desperate maneuver involving a diacritic feature. But unlike the previous case there is no way of handling the change in terms of non-application of rules. With the theory that we have so far the best we can do is to say that the change consists of two separate but simultaneous events: first, the rendering inapplicable of (6), and second, the entering of (3) instead. That is, we are forced to treat this

event as a composite product of two simultaneous changes, one of which alone would have far more spectacular consequences than the two have together. This leaves us completely in the dark as to why so many dialects (quite independently of each other, as is clear from the geographical distribution) should have undergone such a complicated change.

We see that in order to account for such examples, we must relax the proposed restrictions to the extent of allowing a single historical event to involve two distinct processes. But if this is done, then at the same time the gates are thrown wide open for innumerable absurd descriptions in a host of other cases. For example, going back to Ex. 1, there are now several other alternatives which add to the general arbitrariness: the change might consist in simultaneously making (3) inapplicable and adding an exact replica of it after (2), or in simultaneously making (2) inapplicable and adding an exact replica of it before (3).

Also, it is now just as easy to express the reverse change, i.e. a change as a result of which the order (2, 3) changes into (3, 2), so that all umlauted  $\bar{x}$ -segments which are derived by way of  $\bar{a}$  from  $\underline{at}$  would revert to their intermediate representation  $\bar{a}$ , while the  $\bar{x}$ -segments derived from  $\bar{a}$  would stay unchanged. There is no doubt that a theory of linguistic change should either completely exclude the possibility of such a change or at least reflect in some way the obvious fact that it would be a far more complex and unlikely historical event than what actually happened. But as we have been forced to relax and extend the theory by the brute necessity of somehow accommodating one set of data we have lost the capacity of expressing in it the empirical facts that originally motivated it.

There is another consideration, of a somewhat different kind, which shows that changes such as Ex. 1-3 are at bottom very different from the examples of Ch. 1. A characteristic fact about the latter - the "ordinary" sound changes - is that they can spread over large areas with astonishing rapidity. Linguistic borders present no obstacle to the diffusion of sound change (Jakobson, 1931). But changes such as Ex. 1-3 have quite different diffusion properties. They rather tend to spread slowly and take over gradually, with the original and changed grammars often coexisting for centuries as related dialects or styles in a way that is quite unusual in sound change. When we can follow their origin in detail it frequently turns out that they develop independently in parallel ways in many dialects at once. In some respects, we might say, the difference resembles the difference between the spread of contagious diseases and congenital diseases. There is a good reason for this resemblance, to which I return below.

At any rate, Ex. 1-3 evidently represents a new kind of phenomenon. In the next section of this chapter I will try to work out in somewhat more detail its special properties.

## 2.2. Loss and reordering

Examples such as those in section 2.1 fall into two types depending on the nature of the relationship between the original and the changed grammar. In one type, represented by Ex. 1, the changed grammar has the same rules as the old one, but in a different order. In the other, the changed grammar lacks a rule (Ex. 2) or part of a rule (Ex. 3) of the original grammar.

Matthews (forthcoming) has found many cases of the latter type in the recent history of the Siouan languages, and has proposed that they represent a special type of linguistic change, in which rules are generalized. I will refer to it here as loss. In the following I will look at loss first as it affects the output of the grammar, and then as it affects the grammar itself.

Recalling the bipartition of rules into the structural analysis (SA) and the structural change (SC) that was made use of in section 1.6, consider the way the effect of loss depends on where the loss takes place. The basic distinction that I want to draw is between loss that affects the SC ("SC loss"), regardless of whether it affects the SA as well, and loss that affects the SA only ("SA loss"). Thus I regard loss of an entire rule or subrule, as well as a case like Ex. 3, as instances of SC loss.

When there is loss in the SC of a rule, as in both Ex. 2 and Ex. 3, the effect must be that an existing alternation is simplified, i.e. allomorphic variation decreases. Thus, in Ex. 3 such loss results in the replacement of a two-feature alternation kraft:kreftig with a one-feature alternation kraft:kräftig. The extreme case where an entire rule is lost is reflected in the output as the complete elimination of an alternation, as in Ex. 2 where the word-final devoicing rule is lost.

When there is loss in the SA but not in the SC, the output is affected in a different way. Here existing alternations are extended to further cases, i.e. allomorphic variation increases. Suppose, for example, that in some dialect umlaut is restricted to non-compact vowels:

$$(7) \quad [-\text{consonantal}] \rightarrow [-\text{grave}] / \dots \\ [-\text{compact}]$$

In this hypothetical dialect we would have umlaut in hof:ho<sup>f</sup>fe but not in kraft:krafte. Loss of the specification [-compact] from the SA of (7) would have the effect of extending umlaut to a set of cases like krafte which were not affected before. A somewhat more complex actual example is the following.

Ex. 4. In Old English (Luick 1921, §204) vowels were shortened before clusters of three consonants (e.g. gōdspell > godspell, brāmblas > bræmblas) and before two consonants in the third syllable or further from the end (e.g. bliédsian > bledsian). In isolated cases the shortening also took place before geminate consonants. The OE shortening rule was

$$(8) \quad V \rightarrow [-\text{tense}] / \_ CC \left\{ \begin{array}{l} C \\ \dots V \dots V \end{array} \right.$$

with some further special provision for the isolated shortenings before geminates.

In Early Middle English the shortening is extended to a large group of new cases. As Luick (1921, §352-3) points out, we have to do with a further development of the OE shortenings rather than with an entirely new EME sound change. Shortening now takes place even before just two consonants (e.g. képte > kepte, düst > dust, mëtte > mette) and even before a single consonant in the third syllable or further from the end (e.g. sǖerne > sūerne, ärende > arende). We see that this is all the result of a single change which consists in the loss of one of the required consonants from the environment. The resulting ME (and current English) rule is as follows:

$$(9) \quad V \rightarrow [-\text{tense}] / \_ C \left\{ \begin{array}{l} C \\ \dots V \dots V \end{array} \right.$$

For the current English details, see Chomsky and Halle (forthcoming).

There is another difference, which is particularly important for historical reconstruction, between SC loss and SA loss. Only in the latter are all forms affected by the older form of the rule compatible with the changed form. Hence it is possible to reconstruct SC loss but not SA loss on comparative or even internal evidence if relic forms that bear witness of the form of the older SC survive outside of the productive system of derived forms. In both Ex. 2 and Ex. 3 we were able to make use of such relic forms, but clearly none

could in principle exist in a case of SA loss like Ex. 4. Hence SA loss is in general much more difficult to recover by reconstruction.

In terms of its effect on the grammar, the salient feature of loss is that it results in simplification of the grammar according to a notion of simplicity implicit in most of traditional work in linguistics, which the notational conventions of generative grammar are intended to explicate by translating complexity into length (Halle, 1961, 1962; Chomsky, 1965). The question of how this is to be done remains moot in many points, but not in ways that would affect the conclusion that in these and numerous other examples where the form of a rule changes, the outcome is simpler than the original.

We have very little empirical knowledge of how, in practice, such loss takes place. Two of the special properties of this kind of change, its simplificatory character and its slow diffusion, strongly suggest that its origin must be sought not in mature speakers but rather in the process of language acquisition. In particular, the following way in which loss might arise suggests itself (Matthews, forthcoming, cf. also Stampe, forthcoming).

Suppose that the child has constructed the optimal, "simplest" grammar for a certain body of linguistic experience. Further data that he encounters may subsequently motivate a reanalysis. If no reanalysis is made, the grammar will "congeal" as the child matures and passes the stage where complex unconscious skills are easy to acquire. That no reanalysis is made may be due to the fact that the language in fact has no counter-data. In this case the child has arrived at the "right" grammar. But - and this is of vital importance - that is not the only possible reason why a reanalysis may fail to take place. It may well be that the language has counter-data which do not register on the child in the language-acquisition stage, perhaps because of their rarity, the child's limited linguistic experience, some mental or perceptual limitations of the child, or probably most commonly just because of the inevitably fragmentary and incomplete nature of the data at the child's disposal. In that case, linguistic change has taken place: the child arrives at the "wrong" grammar, i.e. not the grammar of those whose speech provided his linguistic experience.

To turn to Ex. 3 for concrete illustration of this "fortunate ignorance" theory, suppose that the child's linguistic experience includes examples for umlauts of o, u etc., but that examples for the umlaut of a to e such as (5) are missing by chance from this corpus. The simplest and most general rule compatible with this curtailed

linguistic experience is just the changed rule (3). One could suppose, in this particular example at least, that most children learning the original dialect will at some early point have this rule in their grammar but will later reanalyze it to (6) as forms like (5) appear. Historical change, according to this theory, takes place when, for some reason, such post-editing of the grammar is neglected.

The notion of fortuitous ignorance in language-learning (which, incidentally, has a long tradition) explains why the directionality of these changes should be determined by simplicity. If the wrong grammar is reached in fortuitous ignorance of the counter-data, we should expect the directionality of such imperfect learning to reflect, as it does, the evaluation measure used by the child in selecting one of the alternative grammars compatible with his linguistic experience. Note that it would be perfectly compatible with the fortuitous ignorance theory that imperfect learning should result in a changed grammar which is not simpler but equally simple than the original grammar; it is merely necessary that imperfect learning from fortuitous ignorance should not increase complexity.

In many cases, however, it is implausible to ascribe loss to fortuitous gaps in the data. We must rather suppose that the child can actually override the data and arrive at a simpler grammar than his linguistic experience warrants. Otherwise it is difficult to see e.g. how the German word-final devoicing rule could have been lost (Ex. 2) or the OE shortening rule could have been simplified (Ex. 4). In the latter it is striking that the original rule applies to a very few cases (because the required environment was rare) but the simplified one cuts right across the phonology creating the productive vowel alternations so characteristic of modern English. Later (Ex. 5) I will discuss a case where it is in principle necessary to assume that certain data has been used in constructing a grammar but nevertheless is not present in unchanged form in the output of this grammar. This possibility has long been known, and lexical contaminations and similar types of change studied especially by neo-Humboldtian linguists with their emphasis on the creative aspects of language are typical cases in point. It is nevertheless surprising to see this happen in the very different realm of productive phonological rules. These facts are particularly striking justification for Chomsky's observation that "the structure of the grammar internalized by the learner may be, to a presently quite unexpected degree, a reflection of the general character of his learning capacity rather than the particular course . of his experience" (1964, 113).

It seems likely that both fortuitous ignorance and overriding of data contribute to imperfect learning. To decide, in a particular

case, which is involved, two factors must evidently be considered:

1. the magnitude of the effect on the output
2. the magnitude of the simplification in the grammar.

The greater the effect on the output (in terms of the number and frequency of altered forms, etc.), the unlikelier it obviously is that the change is due to fortuitous ignorance. In the case of relatively gross deviations such as those in Ex. 1-4, it is surely out of the question that the counter-data has simply by chance not appeared in the linguistic experience of children, and we must rather assume that the simplification has, for some obscure reason, been "worth" the violation of observational adequacy that the child has committed. On the other hand, the smaller the simplification which results from the change, the unlikelier it is that the change is due to overriding of data. It is hard to see why a child should fail to match its linguistic experience if no compensating advantage accrues. Perhaps, in particular cases, there might be a question of linguistic play, or there might be a very deep regularity in the language which, though effecting a great simplification in the grammar, is very hard to discover because of its abstract nature and hence might be missed, leading to a grammar which is more complex and generates the wrong language to boot. These would be very special situations, however, and I shall assume that imperfect learning cannot, in general, increase complexity, and that cases where there is no change in complexity must necessarily be due to fortuitous ignorance.

In summary, then, imperfect learning can take place in two ways, mainly: by fortuitous ignorance, typically resulting in small changes with little or no simplification of the grammar, and overriding of data, typically resulting in large changes with considerable simplification in the grammar.

In the remainder of this section and in section 2.3 I will take up the second type of change, reordering. I will begin with some further examples of this process.

Ex. 5. In his Development of the Canaanite Dialects (1939), Z. Harris describes the following events. By a very early Semitic sound change, diphthongs were monophthongized:

$$(10) \quad \left\{ \begin{array}{l} \text{ay} \\ \text{aw} \end{array} \right\} \rightarrow \left\{ \begin{array}{l} \bar{e} \\ \bar{o} \end{array} \right\}$$

Some centuries later, a rule

(11)  $h \rightarrow \emptyset / V \_ V$

was added to several of the dialects. For example, h was lost in the suffix -ahū, thus giving rise to a new diphthong in this suffix. By a convincing orthographic argument Harris is able to determine that at one time the diphthong -aw existed in pronunciation in this suffix. But quite soon this new diphthong -aw in turn also became -ō.

We would not want to say that this was some independent new sound change, unrelated to the previous operation of (10). To propose (as Harris does on pp. 31, 56) explanations in terms of preservation of the syllabic structure and so forth would be to overlook the fact that we must not only explain why the new diphthong was eliminated, but also why it was eliminated in the same way as the old one, becoming ō and not, say, ā. It must evidently be due to the fact that the dialects at this point contained the morphophonemic rule (10). Evidently, then, there was a change from the order (10, 11) which underlies the stage with the suffix -aw, to the order (11, 10) which is the grammar after the suffix becomes -ō.

Ex. 6. The development of the Slavic palatalizations offers several interesting examples of reordering, of which I will cite the simplest here.<sup>6</sup> By the so-called First Palatalization, k and g become č and ž, respectively, before front vowels and y, e.g. \*kito > cito 'what', \*givu > živu 'alive'.

(12a)  $\begin{bmatrix} [-\text{consonantal}] \\ [-\text{diffuse}] \end{bmatrix} \rightarrow \begin{bmatrix} [-\text{grave}] \\ [+ \text{strident}] \end{bmatrix} / \_ \begin{bmatrix} [-\text{consonantal}] \\ [-\text{grave}] \end{bmatrix}$

But the resulting voiced affricate ž has become a continuant ž in all Slavic languages by the rule

(12b)  $\begin{bmatrix} [+ \text{voiced}] \\ [-\text{grave}] \\ [+ \text{strident}] \end{bmatrix} \rightarrow [+ \text{continuant}]$

e.g. \*živu > živu.

Subsequently new front vowels came to stand after velars by the rule

(12c)  $ai + \check{e}$

By the so-called Second Palatalization, k<sub>1</sub> and g<sub>1</sub> (derived from k and g by an earlier rule) became c and ž before these front vowels, e.g.

\*kļena > cěna 'price', \*gļelo > žēlo 'very':

(12d)	$\begin{bmatrix} +obstruent \\ -grave \\ -strident \\ -diffuse \end{bmatrix}$	$\rightarrow$	$\begin{bmatrix} +strident \\ +diffuse \end{bmatrix}$
-------	---	---------------	---

The resulting affricate ž, unlike the earlier ž, is retained in Old Church Slavic and in modern Polish. The grammars of these languages have (12a-d) as phonological rules in an order that matches their relative chronology. But elsewhere in Slavic, ž also has been replaced by its corresponding continuant, namely z, e.g. žēlo > zēlo. These languages have the same four rules, but (12b) must here follow (12d) in order to apply to the affricate produced by the second palatalization as well. I propose that the historical change by which this latter grammar developed is precisely the shifting down of (12b), rather than the introduction of some entirely new rule.

But I have still not explained why the reordering should have a specific directionality. Apparently there is an asymmetry of rule order types: a particular type of order is unstable and may get inverted, i.e. with certain pairs of rules A, B the order

1. A
2. B

may get inverted but the order

1. B
2. A

may not. It is important to realize that there is no necessity for the reordering to take place; we can merely delimit the direction it will have if it takes place. This is not different from the fact that we cannot predict what specific sound changes will take place but merely delimit a set of possible sound changes. "Il en est comme de l'eau de pluie qui doit prendre un chemin prévu (gouttières, égouts, conduits) une fois qu'il pleut. Mais la pluie n'est pas une nécessité." (Kuryłowicz, 1949). The problem is how and why such order change takes place, and in particular under what conditions such asymmetry of order obtains. To discuss this it is first necessary to consider in a general way some aspects of the role of rule order in generative grammar.

In any component of a grammar, some but (in general) not all rules are functionally ordered with respect to each other. For example,

the Am. English rule that  $\check{ɔ} \rightarrow \bar{a}$  must precede the rule that vowels are lengthened before voiced consonants (lest /pɔd/ become [pɔ̄d]), but neither of these must necessarily precede or follow the rule that initial voiceless stops are aspirated. Imposing a fixed order on all rules necessitates an arbitrary decision in cases such as these where no fixed order is internally motivated. It forces us to treat alike cases where order matters and cases where it does not. As an alternative, one might consider a notation in which there would be a motivated way of omitting order indications where no specific order can be justified. For example, if we had a notation in which fixed ordering increases complexity, the evaluation measure will then force us to indicate order only where it has linguistic significance. One way in which this could be done is the following.

Within a component of the grammar, the rules are assigned integers 1, 2 ... n. Ordering is indicated by associating with each rule a complex feature [precedes K], where K is a set of integers  $< n$ , possibly null, in a component with n rules. Thus, to indicate that rule 1 must apply before rule 2, rule 2 is marked [precedes 2]. This complex feature is transitive, i.e. if we have rules 1, 2, and 3 where 1 is marked [precedes 2] and 2 is marked [precedes 3], then also automatically 1 is marked [precedes 3]. It is important that the feature [precedes K] is an integral part of the grammar and hence figures in the determination of relative simplicity of grammars. On the other hand, the order in which rules are listed has no systematic status in the proposed notation, although it is naturally convenient to list rules in the order of their application.

With the notation we have so far, the evaluation measure assigns equal simplicity to the following two sets of three ordered rules:

- (13) 1.  $a \rightarrow b$       [precedes 2]  
 2.  $b \rightarrow c$       [precedes 3]  
 3.  $c \rightarrow d$
  
- (14) 1.  $c \rightarrow d$       [precedes 2]  
 2.  $b \rightarrow c$       [precedes 3]  
 3.  $a \rightarrow b$

(13) and (14) are both regarded as less simple than the following set of three unordered rules:

- (15) 1.  $a \rightarrow b$   
 2.  $c \rightarrow d$   
 3.  $e \rightarrow f$

I would now like to propose a further modification. A particular type of fixed order, unmarked order, of which (13) is a typical example, is inherently simpler than its inverse, marked order, exemplified in (14). I suggest that the notation of (13) be changed as follows:

- (13') 1.  $a \rightarrow b$
- 2.  $b \rightarrow c$
- 3.  $c \rightarrow d$

Thus (13') equals (15) in simplicity and both are simpler than (14). The correct order of application will be assigned to (13') by a general principle whose form is discussed in the next section.

With regard to the evaluation measure I suggest that there is no trading relationship between complexity of phonological feature specifications and complexity of rule ordering. Rather, the feature specifications always take precedence over the ordering in evaluating grammars. Thus, we pick the grammars with the least feature specifications, and among these the one with the least order specifications, from among alternative proposed solutions.

Thus the asymmetry of rule order types is formally reflected in that rules will automatically be placed in unmarked order unless there is a specific marking to the contrary. This is exactly analogous to the Pragian notion of markedness according to which certain binary phonological and morphological features are asymmetric in that their one term is general, simple or "unmarked" and the other is restricted, complex or "marked" (Jakobson 1932). The natural way to incorporate this insight in the framework of generative grammar is to indicate only the marked term of each feature in the underlying representation and to introduce the other by general rules when there is no specific marking to the contrary. For example, phonological segments are automatically specified as unaspirated unless they are expressly marked as aspirated. (Cf. Chomsky and Halle, forthcoming).

The reorderings of phonological rules discussed above were changes from marked to unmarked order. Reordering, then, is just a special case of loss, namely loss in the complex feature that marks order. Evidently, also, reordering originates in language acquisition in the manner described for loss above.

If the above proposals are to have any significance, they must be accompanied by a general criterion for determining rule order markedness. In the next section I will examine several possible approaches to such a criterion.

### 2.3 Unmarked order

An interesting proposal, due to T. G. Bever, is that asymmetry of order is determined by the relative simplicity of the rules, in such a way that the simplest rules will come last in unmarked order. Underlying this suggestion is the idea that maximum symmetry of the output would be achieved by applying the simplest, most general rules after the more restricted rules. This condition could easily be made precise and would, if correct, give a simple and straightforward characterization of rule order markedness in terms of independently motivated theoretical concepts.

But this condition can at best provide a crude approximation of our goal, and there are many cases in which it yields the wrong result. Suppose, indeed, that we have a pair of rules (A, B) where the unmarked order is as given and B is indeed the simpler rule. Clearly we can increase the complexity of B in numerous irrelevant ways without changing the markedness of the ordering in the slightest, e.g. by adding to B some long list of idiosyncratic environments. Thus, in Ex. 1 the markedness would be no different if the Umlaut rule were more complicated than Monophthongization - say, if umlaut applied only before certain peculiar consonant clusters and in certain grammatical categories. Actual examples where an unmarked order requires precedence of a simpler rule in contradiction with the condition under discussion are not hard to find. This is the case e.g. in Ex. 5, when the relevant rules are stated in feature notation.

Perhaps, then, it is not the gross relative simplicity of rules but some more complex property of their configuration that determines their asymmetry of ordering. We notice, for example, that in one characteristic type of unmarked rule sequence (e.g. Ex. 6) the first rule introduces representations that satisfy the SA of the second rule:

1. [ ] → [ αF ]
2. [ αF ] → [ ]

The question is whether such observations can be extended and systematized to provide an adequate characterization of order markedness in terms of the partial recurrence of feature specifications in rules. Consideration of the more complex cases shows, I think, that this is impossible. It is difficult to see, for example, what intrinsic properties of the rule configuration determine the markedness of the order in such cases as Ex. 1 or Ex. 5.

Aside from particular counter-examples there is a general reason why both the above conditions should fail to characterize order markedness correctly. This is that both are based solely on the configuration of rules and do not take into consideration the relationship of the rules to the representations to which they apply. Though it may be convenient to ignore this, it is clear that it must be included in some way when the order types are defined. It is only in terms of the total grammar that the notion of relative simplicity has significance, and relative simplicity of rule ordering is no exception to this.

Consider, for example, a hypothetical language which (like English) has a rule (A) that vowels which agree in rounding and gravity (e.g. e but not a) undergo some change P. Suppose that this language also has a rule (B) that makes compact vowels nongrave in an environment Q. It is clear that the situation is radically different depending on whether the grave compact vowel to which (B) will apply is ɔ or a. If it is a, then (B) creates a new vowel ä which fits the SA of (A). In this case the unmarked order is clearly (B, A). But if the grave compact vowel is ɔ, (B), were it to apply first, would on the contrary remove a vowel which fits the SA of (A). There is some evidence, which I will present below, that in this case just the opposite order (A, B) is unmarked.

Examples such as these suggest that the approach in terms of rule configurations is fundamentally misguided. A different approach is the following, which involves Chomsky's concept of intrinsic order.<sup>7</sup> Suppose that the rules are applied in free ("random sequential") order, i.e. we run through the rules in arbitrary sequence, making the convention that the derivation is terminated when a representation is obtained to which none of the rules apply. If it is the case that each input determines just one unique output, which is the same no matter what order of application we may happen to pick, then the intrinsic order is that fixed or partially fixed order which yields just this output. For example, if we have the rules.

1.  $a \rightarrow b$
2.  $b \rightarrow c$

then application of these rules in free order to input strings of a's and b's will always result in strings of c's of the same length. The fixed order which has this property is the one given above, i.e. (1,2). This is thus the intrinsic order of these two rules. Intrinsic order is always unmarked. For example, the output of the changed but not the original grammars of Ex. 1, 5, 6 would result from

application, to the correct underlying forms, of the correct rules in intrinsic order. Note that it does not fail in a case like Ex. 5, where the approach in terms of rule configurations does fail. Thus we might consider defining unmarked order as intrinsic order.

Unmarked order, so defined, has several other properties as a consequence. Thus, phonological rules in unmarked order are phonetically true in the sense that each rule, interpreted as a condition on the phonetic output, is satisfied (I owe this point to R. Stanley). Also, this definition guarantees that rules in unmarked order have maximal applicability in that each rule applies functionally in the maximum of possible cases (i.e. vacuous applications of rules are minimized).

A definition of unmarked order as intrinsic order in this sense is undoubtedly correct so far as it goes. That is, in no case that I know of does it go wrong by assigning unmarked status to a marked order. In certain cases, however, it appears that it does not go far enough and conversely leaves an unmarked order marked. It does not provide for asymmetry where for some input different ways of applying the rules in free order yield different outputs. In such a case, there is no intrinsic order and we must consider the order marked both ways, under this definition. But there are indications that even here an asymmetry of marking may obtain. In support of this I would like to cite three further examples and then propose a definition by which order markedness can be extended to include these new cases without changing the outcome in the old cases.

Ex. 7. In the dialects of Northeastern Switzerland the back mid vowel o becomes lowered to ɔ if it immediately precedes a nongrave (dental or palatal) true consonant or r. Cf. in the Kanton of Schaffhausen (Wanner, 1941):

Retention of o:

- before 1: foll, holts, gold
- before labials: grob, ops, hobəl, xnopf, dobə, ofə, xopf
- before velars: xoxxə, knoxxə, rokx, kflogə, bogə.

Lowering to ɔ:

- before r: hɔrn, tɔrn, ſɔm
- before dentals and palatals: rɔss, xrɔtt, lɔsɔ, ksɔt̩ə, bɔdə, pɔʃt.

The correct distribution of allophones is given by the rule

(16)	$\begin{bmatrix} \text{-consonantal} \\ \text{-diffuse} \\ \text{+grave} \end{bmatrix}$	$\rightarrow$	$\begin{bmatrix} \text{+compact} \\ \text{-grave} \\ \text{-lateral} \end{bmatrix}$
------	---	---------------	---

It is imperative that (16) be restricted to the back vowel o. The umlauted variant ö of this vowel is not lowered. The plurals of bog and bod are böge and böde, both with a noncompact ö. Hence the lowering rule (16) must follow the Umlaut rule in this dialect. This is the situation in many marginal dialects of NE Switzerland. The more central dialects show a different state of affairs.

I will take a dialect which in all other relevant respects is identical to that of the Schaffhausen area, namely that of Kesswil in neighboring Oberthurgau (Enderlin, 1913). Rule (16) operates in unmodified form in this dialect. All the vocabulary items cited above for the Schaffhausen dialect are also found, with the same distribution of o and ö, in Kesswil. But the difference is that Kesswil, along with most of NE Switzerland, has ö as the umlauted form of o but ɔ as the umlauted form of ɔ. In these dialects the plural of bog is böge but the plural of bod is böde.

One solution would seem to be that the lowering rule was simplified to apply to rounded vowels regardless of gravity. But this fails since phonemic ö does not lower to ɔ in the environment of (16). The crucial cases are such forms as plötsli and fröss 'frog' (orig. a plural form). The behavior of these isolated cases shows conclusively that we are in reality not dealing with a lowering of ö to ɔ at all, but rather with the umlaunting of ɔ as well as o. That is, umlaunting now takes place after rather than before (16).

Ex. 8. A very widespread sound change in German dialects (Schirmunski, 1961, 212) is the rounding of ä to ɔ. As ä, the umlaut of ä, is unaffected by this change, it brings about alternations between ɔ and ä like the following:

<u>šwɔn</u>	'swan'	$\sim$	<u>šwän</u>	'Pl.'
<u>špɔt</u>	'late'	$\sim$	<u>špäter</u>	'later'

Because of its umlaut and for other reasons, phonetic ɔ is morphophonemically represented as ä and the rounding is added by the rule

$$(17) \quad \overline{a} \rightarrow \overline{\circ}$$

which must apply after Umlaut since the umlauted vowel is unrounded. Many modern German dialects have just this system, e.g. the IG dialect

of Bleckede (Rabeler, 1911) or the Swiss dialect of Mutten (Hotzenköcherle, 1935). In others, such as the Swiss dialect of Schaffhausen (Wanner, 1941) the system has changed in that the umlauted form of ɔ is ö, e.g. swɔ̄n, spɔ̄ster. The grammar has changed in that umlaut now applies after rather than before (17). As phonemic ä in words like tsä 'tough', lǟr 'empty' stays unrounded it is clear that the possibility of a generalization of the rounding rule to all long compact vowels is excluded and we are indeed again faced with a case of reordering.

Ex. 9. A third, very similar example concerns the order of two rules pertaining to voiced obstruents in German. One of them, which is historically the older of them is the devoicing of obstruents in final position (e.g. tāg > tāk) that was already discussed in connection with Ex. 2. The other, found only in a certain group of dialects (Schirmunski, 1961, 302 ff.) is the spirantization of postvocalic voiced obstruents, e.g. tāgə > tāvə, sāgt > sāyt (> sāxt). Originally, devoicing preceded postvocalic spirantization. Since with this order morphophonemic final voiced stops lost their voicing before spirantization applied they remained stops and we had tāk but tāvə. Some Alsatian, Bavarian and Middle German dialects still have this system. More frequently, the reverse ordering is found, with final voiced stops undergoing first spirantization (tāg > tāv) and then devoicing (tāv > tāx). This order is typical of the Low German dialects. We know that this order is a secondary development because some words like weg where the voicing of the stop had no morphophonemic support failed to spirantize even in the reordering dialects. This would be inexplicable unless we suppose that the devoicing was historically earlier even in these dialects in spite of the fact that it is synchronically later.

Let us call cases like Ex. 7-9 Type II reorderings, to distinguish them from the other cases, Type I reorderings. It is clear that both are fundamentally of the same simplificatory character. But the notion of free order or intrinsic order cannot by itself be used to define markedness in Type II reorderings, as was pointed out above. However, the asymmetry of rule order will extend to these new cases if we assume that maximal applicability itself is the defining property of unmarked order. This can be made more precise in the following way, suggested by Matthews. As before, only marked order is indicated in the grammar. Where no order is indicated, the rules are applied in free order and that output which results from the longest derivation is selected. The proper measure of length is the number of distinct representations in the derivation, as we do not want vacuous applications of rules to increase the length of a derivation. For well-

formedness it must be required that order be marked both ways when there are derivations of equal length that yield different outputs from the same underlying representations.

An argument for extending the asymmetry of order in this way is that the addition of Type II reorderings creates a remarkable parallelism between reordering and loss. I already distinguished between loss which affects the SC and loss restricted to the SA of rules and noted two characteristic differences between their respective effects on the output. Precisely these differences also recur between the two kinds of reordering. Type I reorderings correspond to SA loss (extension of alternations, absence of relic forms); Type II reorderings correspond to SC loss (simplification of alternations, presence of relic forms).

The question of the exact characterization of order markedness is obviously best left open for the time being. On the one hand, it is necessary to seek more empirical evidence, not only from historical change but also from language acquisition and language disorders. We would, for example, expect a progression from unmarked to marked order in language acquisition and in recovery from aphasia. On the other hand, the form of ordering in grammatical rules needs more study both from the viewpoint of descriptive adequacy and from the viewpoint of its formal properties.

It is an interesting psychological question why one type of order should be simpler than another. The rules of a language constitute a system that the child somehow discovers on the basis of a limited linguistic experience, somehow commits to memory and then somehow puts to continual use in producing and understanding speech. All of this activity normally takes place on a completely unconscious level. The ultimate reason for why grammar A is simpler than grammar B must be that a human being is so equipped that he can more easily discover, remember or use grammar A than grammar B. We would expect that the notion of unmarked order eventually will tie in with such considerations. At present, there is little that we can say with certainty about questions of parole, but it is reasonable to suppose that relative ease in language acquisition will somehow involve the relative "surface accessibility" or "discoverability" of linguistic generalization, relative ease of internal representations will involve considerations of long-term memory, and relative ease in speech activity will have to do with short-term memory ("computing space"). Along these lines one would hope to find the reason why such factors as phonetic truth, maximal applicability and intrinsic order apparently contribute to the simplicity of a grammar.

## 2.4 Proportional analogy

The traditional explanations for the kind of phonological change I have discussed in this chapter are quite diverse. I want to discuss here the adequacy of the most commonly encountered one, namely analogy. The proposal that the changes illustrated above are due to analogy involves both a grouping with certain other kinds of change and a theory to account for such changes. Thus, it is one thing to claim that Ex. 1 illustrates the same mechanism as e.g. the change from brethren to brothers and quite another to claim that the mechanism in both cases is the mutual proportional influence of allo-morphs. It is obviously possible to accept the first while denying the second. In fact, that is precisely the position I would like to defend below.

It is indubitable that loss and reordering in phonological rules resemble the regularizations of paradigms like brethren > brothers that serve as the stock examples of analogy. In both cases we have to do with change of grammars by partial modification rather than by the introduction of new rules.

In both cases the process involves a simplification of the grammar. It is obvious why this is true for the morphological examples. The regularizing of the plural of brethren is in fact just a trivial case of loss in that some special rule such as brother + Pl → brethren is dropped from the language. A change like that of knelt to kneeled is somewhat different since here the rule itself survives (cf. felt, dealt) but the verb kneel is removed from its domain by loss of the special marking associated with kneel that indicates strong inflection.

Since there is a system of rules underlying the inflection of strong verbs it is not necessarily a complication of the grammar if a verb shifts in the reverse direction from weak to strong conjugation. In verb stems of a certain phonological form (in English, perhaps verbs in -ing) it may well be weak inflection which is exceptional and has to be indicated by a special mark in the dictionary, the loss of this mark making the verb strong.

Support for the identity of morphological analogy with loss and reordering also comes from the fact that their diffusion properties are strikingly similar. As noted in 2.1, loss and reordering usually do not sweep rapidly across large areas as ordinary sound changes often do, but rather spread slowly with frequent parallel independent development. But this is very characteristic of analogical changes in morphology. In both cases this is best explained by assuming such changes to be less susceptible to borrowing by mature speakers.

For these reasons I agree with the traditional view that there is a single form of linguistic change at work in cases of this type (Sapir, 1915 255; Schuchardt, 1885, 49, 61; Schirmunski, 1961, 199 and many other references). If this is correct, then any proposed theory of analogy must be able to account for the phenomena of loss and reordering in phonological rules. Below I would like to mention some of the ways in which the proportional theory of analogy is inadequate in this respect.

I take the proportional theory of analogy to be the view that individual linguistic forms may change under the influence of other individual linguistic forms according to a certain proportional pattern.<sup>9</sup> Thus, the forms kneeled is explained, typically, as the result of solving for X, in some way, a proportion such as peel:peeled = kneel:X.

This is all quite empty, obviously, unless some restrictions are put (1) on the kinds of forms that may be linked to such a proportion and (2) on the kind of influence that the forms may exert on each other. Let us consider quite briefly each of these problems.

It is clearly not enough to require simply that in a proportion  $A:B = C:D$ , A, C and B, D belong to the same grammatical category. We do not want to allow a proportion walk:wander = talk:X resulting in a new verb tonder 'chat'. One of the few who have recognized the need to raise the proportional theory of analogy from utter vacuity by restricting the structure of proportions is J. Kuryłowicz. Kuryłowicz (1949) has put forth six extremely interesting Laws governing the process of analogical change, of which the second bears directly on the point I just made. Its original formulation is as follow:

"Les actions dites 'analogiques' suivent la direction: formes de fondation → formes fondées, dont le rapport découle de leurs sphères d'emploi."

This principle correctly rules out the absurd proportion cited above, but unfortunately leads to trouble in another direction. It wrongly excludes the possibility of back-formations like that which led to the introduction of the simplex verb edit on the basis of forms like editor, edition, in the English of a few centuries ago. The view of analogy as a simplificatory process of imperfect learning provides a straightforward account of such back-formations and also an explanation of their special character. For syntactic reasons (e.g. the existence of NP's like the editor of this book) it was necessary, even when there was no actual form edit, to postulate just

such an underlying form from which editor, edition had to be obligatorily derived (Chomsky, 1965). The analogical change is simply that the anomalous obligatory derivation becomes a normal optional derivation. But to explain this and similar back-formations within a proportional model of analogy, it is necessary to allow "inverted" proportions with the derived forms (formes fondées) on the left, such as writer:write = editor:edit, in clear violation of Kuryłowicz's second Law of analogy. However, it is possible that his third Law may allow the required inverted proportion to hold:

"Une structure consistant en membre constitutif plus membre subordonné forme le fondement du membre constitutif isolé mais isofonctionnel."

The reason I am not sure whether this is so is that the terms of the proportion must be "isofunctional" and the concept "function" is taken for granted by Kuryłowicz in a way that makes it impossible for me to determine in this particular example whether the requirement of isofunctionality is met or not. A great deal hinges on this requirement, since if it were dropped from the third Law, the third and second Laws would in most cases assign the opposite proportion and thus would combine to make each other largely vacuous, as far as I can see.

In general, Kuryłowicz's Laws are better regarded as, perhaps, statistical tendencies of some kind rather than as laws in the usual sense. Thus, according to the first Law,

"Un morphème biparti tend à s'assimiler un morphème isofonctionnel consistant uniquement en un des deux éléments, c.-à-d. le morphème composé remplace le morphème simple."

Kuryłowicz illustrates this with the case of German umlaut, which has spread analogically to originally unumlauted plurals of former a-stems, e.g. MHG boume > Bäume. Thus the "bipartite morpheme" consisting of umlaut plus the ending has spread at the expense of the "simple morpheme" consisting of the ending alone. But in the following quite similar case just the opposite kind of development has taken place. The Germanic strong past participles were in part doubly characterized by ablaut and the consonant change resulting from Verner's Law, e.g. OE cēasan:coren, and in part (where the stem consonant was not susceptible to Verner's Law) by ablaut alone, e.g. bēodan:boden. In the modern Germanic languages Verner's Law has ceased to operate (except for a few forlorn relics) but ablaut is to a considerable extent retained, e.g. choose:chosen. Thus the "simple

"morpheme", in Kuryłowicz's sense, has in this case ousted the "bipartite morpheme". If we regard analogy as a simplificatory process, both kinds of developments are just what we should expect. Once a rule loses its phonological conditioning environment, the grammar may become simplified in that the rule becomes contingent on some other, independently needed environment (e.g. umlaut is becoming contingent on strong plurals in German) or it is lost altogether. The simplificatory process in the former case may well involve a decrease in the application in the rule as well as an increase, in contrast with phonologically conditioned rules where SA loss always causes an increase. This difference between the analogical development of morphologically and phonologically conditioned rules is another facet of analogical change which is outside of the scope of proportional theories of analogy but follows naturally from the theory proposed here.

Kuryłowicz's fourth (and perhaps most widely accepted) Law is the following:

"Quand à la suite d'une transformation morphologique une forme subit la différenciation, la forme nouvelle correspond à sa fonction primaire (de fondation), la forme ancienne est réservée pour la fonction secondaire (fondée)."

This would seem to conflict with some of the developments of the English strong verbs. Consider cases where strong conjugation has been replaced by weak conjugation in some verb but both past forms still remain in use. It is not at all impossible that just the new, analogical weak form comes to be used in a "secondary function" with the old strong form surviving in the "primary function". Thus, shined is used in one restricted sense (I shined my shoes) but the general past tense (both of the causative and the non-causative) is shone. The past of weave is wove in the primary function but weaved in certain technical uses (evading aircraft, etc.). Similarly, wind is strong in ordinary usage (wound) but weak in the special nautical sense. It seems to me, therefore, better to return to the usual, more modest but also more correct formulation that unproductive (traditionally, "de-etymologized") formations escape analogical influence, a formulation that accounts for many of the facts Kuryłowicz's fourth Law is designed for but does not have the same unhappy side-effects. This traditional principle, moreover, needs no express statement in the present theory as it is a perfectly automatic consequence of the rule-bound conception of analogical processes.

Kuryłowicz's fifth Law,

"Pour rétablir une différence d'ordre central la langue abandonne une différence d'ordre plus marginal."

seems plausible and may well turn out to be correct once lent some substance by a general characterization of marginal versus central differences (perhaps in terms of a hierarchy of universal syntactic features, analogous to the hierarchy of phonological features, cf. Bever, forthcoming).<sup>10</sup> On the other hand, the sixth Law,

"Le premier et le second terme d'une proportion appartiennent à l'origine à des systèmes différents: l'un appartient au parler imité, l'autre, au parler imitant."

is so way out as it stands that there must surely be some inadvertent inaccuracy in its formulation. The interpretation which Kuryłowicz's examples suggest is that analogical change requires different coexisting dialects to take place and that analogical change is due to hypercorrection. But if a small child says I bringed it, what hyper-correction could possibly be involved?

At any rate, the first problematic aspect of the formation of proportions that was mentioned above has received at least some attention, largely due to the work of Kuryłowicz. The second, namely the question: what kind of analogical influence may forms exert on each other? has been much more thoroughly ignored from neo-grammatician days up to the present. Clearly there will always be many pseudo-relations between forms that will play no part in the formation of proportions. Thus, on the basis of /day/ : /dayd/ we do not form a new past tense /trady/ from /tray/ by abstracting the wrong principle of past formation "add /d/ after the third segment", or /trday/ ("add /d/ after the first segment that has the same value for the features consonantal and vocalic"), or /trayt/ ("reduplicate the initial segment in final position"). Clearly these are absurdities as each would lead to considerable complication of the conjugational system. It is not so clear, however, how they can be excluded in a proportional model of analogy where such considerations can play no part.

Of more immediate importance to linguistic reconstruction is the fact that between the great mass of obviously correct and obviously false relations there is a group of problematic ones, e.g. the notion "inverted analogy" proposed by van Coetsem (1956). Van Coetsem's theory is that the inherited e → a ablaut of Germanic analogically gave rise to an a → e ablaut. Formally, this amounts to saying that the old rule

(18) [- consonantal] → [- grave] / ...

became, by analogy,

(19)  $\begin{bmatrix} \text{- consonantal} \\ \alpha \text{ grave} \end{bmatrix} \rightarrow [-\alpha \text{ grave}] / \dots$

I am not at all convinced that van Coetsem has found the right solution for this particular problem of Germanic linguistics; the idea of inverted analogy itself, however, makes sense. It is reasonable to suppose that changing fixed polarity values to variably specified values in a rule should make it more general. If that initial assumption is right, then in view of the idea of analogy as a simplificatory process inverted analogy is just what we would expect to find. We should also expect to find unipolar assimilations and dissimilations becoming generalized to the corresponding bipolar processes, e.g. the existence of the process  $\underline{k}\theta \rightarrow \underline{x}\theta$  to lead to a process  $\underline{x}\theta \rightarrow \underline{k}\theta$ . There would be many other diachronic consequences of this kind, which provide an empirical test for the correctness of the theory. However, the view of analogy as simplification, within the theory of generative grammar, also provides a purely synchronic basis for a decision in this and similar moot questions.. That is, we are now in a position to bring problems of diachronic and synchronic linguistics to bear on each other in ways which may be of considerable significance for both.

It is possible that the above criticisms may be met by more adequate formulations of the proportional theory of analogy. I would now like to suggest some respects in which a proportional model would seem to fail beyond possible remedy.

Traditionally, sound change is granted a certain regularity whereas analogical change is considered to be a peculiarly sporadic and irregular kind of process. Actually, the supposed sporadic character of analogy is a mere illusion resulting from the fact that analogy has been systematically studied only in the area of morphology and almost completely ignored in phonology and syntax, where the rule-bound character of analogy is clearly apparent. Cases of phonological analogy such as loss and reordering are typically across-the-board changes that affect the entire vocabulary at once. The reappearance of voicing is complete and not sporadic in Ex. 2, and similarly in most of the other cases. The same generality prevails in "morphological" analogy to the extent that the relevant rules are structured along independent dimensions. Thus, a 3. person singular form commonly generalizes to all singular forms or to all third person forms. It is just where rules lack such structure and their applicability is largely an idiosyncratic property of individual

morphemes, as in English strong verb inflection, that analogical change proceeds morpheme by morpheme in the sporadic fashion that has come to be regarded as typical of analogy in general.

Within the proportional model there is no explanation at all for across-the-board analogical change. It asserts that knelt is replaced by kneeled on the basis of a proportion peel:peeled = kneel:X; and just in the same way that bunt is replaced by bund on the basis of a proportion bunte:bunt = bunde:X. It is a complete mystery in this theory why the former change should affect just this one word whereas in the latter every morphophonemically voiced consonant becomes phonetically voiced in final position in a single historical change. Again, to make any sense out of this kind of change it is necessary to look not just at forms but at the linguistic rules that relate them.

My second general criticism of the proportional theory is that in many indubitable cases of analogical change the allomorphs required to assemble a complete proportion simply do not exist in the language. Such situations will arise when forms that are derived from underlying representations that have no direct phonetic reflex play a role in analogical change. I will discuss two types of cases in which the proportion  $A:B = C:D$  is necessarily defective. First, I will take up cases where the term A is missing, and then cases where the term C is missing.

Typical examples of the former type arise when one of the rules involved in a reordering is context-free. Ex. 5 is a case in point. The following example from Finnish illustrates the same fact even more clearly.

Ex. 10. By an old morphophonemic rule of Finnish underlying mid vowels are diphthongized, e.g. vee > vie. Subsequently, other rules introduce new long mid vowels, e.g. teye > tee. In some Finnish dialects (though not in Standard Finnish) the rules have become reordered so that the new long mid vowels join in the diphthongization, i.e. teye > tee > tie. No proportional explanation for this development seems possible as the underlying form vee is never realized as such. (The grounds on which an underlying vee is set up have to do with morphophonemic rules such as the rule for past formation, e.g. vee + i > vei like saa + i > sai.)

Now consider cases where the right half of the proportion is defective. This will be the case when a form is derived from a base form that does not overtly occur, and by virtue of this derivational relationship undergoes an analogical change. For example, in the Swiss dialect that has lowered the umlaut of o from ö to œ, leaving phonemic

ö unaffected, (Ex. 7), the lowering has also affected a verb like tötterle 'to be scared'. The reason is evidently that this looks like a derived form with the umlauting suffix -lə and is assigned such a structure by the speakers. But the stem is a "cranberry" morpheme that does not occur alone. Any proportion set up to account for the lowering here is necessarily deficient. In the normal case a proportion like bog:bögə = bodə:bɔdə can be set up, but here the right side would simply have no first term. To cite another, entirely similar example, in the dialect from which Ex. 8 is taken, the analogical rounding has also affected the umlaut vowels in forms like xräuel 'pick', f läuwel 'small shovel', yielding xräüwl, f läüwl, as such forms are derived with the common umlauting suffix -əl, although the stems happen to be unique to these words. These examples might easily be multiplied.

In any case, the proportional theory of analogy runs into trouble because it does not allow for the role of rules and underlying representations in analogical change. One might try to modify the proportional theory, accordingly, by allowing underlying representations to figure in proportions and by demanding that proportions be formed in conformity with the grammatical rules of the language. This presupposes some way of knowing what the correct rules and underlying representations are, i.e. it presupposes the entire grammar. But in that case the whole machinery of proportions becomes an entirely irrelevant complication. As I hope to have shown, the grammar of a language is not only a necessary but also a sufficient basis for determining what analogical changes the language may undergo.

## 2.5 Further Types

In this concluding section of the chapter on imperfect learning I would like to take up several other manifestations of this process of linguistic change which are rather more marginal than the two main types of loss and reordering. I will discuss two cases in particular, reordering of redundancy rules, and shifting of rules into cyclical application, which may turn out to have some bearing on moot points of generative phonological theory.

Ex. 11. Literary German (Hochdeutsch) distinguishes three kinds of vowels: tense long vowels, occurring only under stress (e.g. kür 'cure (n.)'); tense short vowels, reflexes of the former in unstressed position (e.g. the first vowel of kurieren 'cure (v.)'); and lax short vowels, which appear both stressed and unstressed (e.g. p Úls 'pulse', p Ulsieren 'pulsate', where laxity is symbolized by capitalization). Assuming that stress and vowel length in German are

either phonemic or introduced by relatively early rules, the situation requires the following two ordered rules:

(20) [α long] → [α tense]

(21) [- stressed] → [- long]

Rule (20) makes long vowels tense and short vowels lax; rule (21) then shortens unstressed long vowels, leaving them still tense, e.g. kürieren > kurieren.

In colloquial speech, however, even among speakers of literary German, the grammar is commonly relaxed to make (20) apply after (21), thus making tenseness contingent on phonetic length rather than underlying length. In such styles there is thus no difference between originally short vowels (e.g. pUlsieren) and secondarily shortened vowels (e.g. kUrieren), both being lax by (20). In view of the fact that speakers tend to lapse into the latter style whenever the speech becomes rapid or colloquial, it seems reasonable to suppose that it is in some sense a simpler system than the former.

The special significance of such an example in terms of the notion of markedness as formalized in generative phonology by Halle and Chomsky is in the fact that (20) is a universal rule which assigns to each segment unmarked for tenseness a value for tenseness that agrees with the value that this segment bears for the feature of length. Ex. 11 appears to suggest that it is possible and in fact simpler to defer application of such redundancy rules until at least some of the phonological rules (P-rules) have applied. This is just what we should expect in view of the fact that the Pragian observations about the appearance of unmarked members of oppositions in neutralized position holds to a large extent for phonetic representations as well as underlying representations. For example, the neutralization of voicing in word-final position in many languages is brought about by a very late rule which, for example, in German must follow vowel apocopation and similar phonological rules. If the asymmetry of markedness were regarded as a property of underlying representations only, we should have no explanation for why there is devoicing and not conversely voicing in final position, and why the sonorants are not devoiced. Given the possibility of deferring application of redundancy rules, we can suppose that the rule is simply that all segments become unmarked for voicing in final position, with a subsequent redundancy rule assigning voicing to sonorants and voicelessness to obstruents.

The possibility of reordering redundancy rules may shed some light on a well-known chapter of West Germanic phonology. (Ex. 12). The front rounded vowels resulting from the operation of the umlaut rule lost their rounding in the majority of German dialects, e.g. tür > tir, löxer > lexer. The unrounding began at the latest in the 12th century in Bavarian and has affected perhaps the majority of German dialects today. The same change took place in early English. As the historical and geographical evidence points to independent development of unrounding in many dialects, Schirmunski (1961, 205) concludes that it was caused by drift. But the drift can hardly be some general language-independent tendency towards unrounding of front vowels, as Schirmunski supposes. Most Uralic and Altaic languages have front rounded vowels, but there is no tendency to unround them here. The drift must rather spring somehow from the particular phonological system of Germanic itself.

In terms of the theory of markedness, both Germanic and Uralic-Altaic (in fact, all languages) have a redundancy rule by which vowels which are unmarked for rounding assume the same value for rounding as they bear for gravity. Thus, back vowels are normally rounded and front vowels are normally unrounded. We see now that the Germanic unrounding consists in deferring the application of this redundancy rule until umlaut has applied, with the result that all umlaut vowels become unrounded. Unrounding resulted from this order change because in Middle High German, when all this happened, the umlaut vowels had not yet become phonemic as in Modern German, but were still derived from underlying back vowels as described in Section 1.2. Now we see also why this change is not found in Uralic and Altaic. In these languages the front rounded vowels are phonemic and hence postponement of the redundancy rule would not have a comparable effect.

These cases, where a principle true on the level of underlying representations is as it were made true on a more superficial level, are by no means isolated. A similar case is described by Halle and Lightner (forthcoming) in connection with the Slavic palatalizations.

Let me go on immediately to another type of imperfect learning which is of some interest. The example to be cited (Ex. 13) is rather similar to Ex. 8 above and concerns the relationship of the same two rules, umlaut and ā → ē (rule 17), in conjunction with a third rule, lengthening of vowels in stressed open syllables. In some Low German dialects (e.g. Rabeler, 1911) they must apply in the order (umlaut, lengthening, rounding) to yield derivations of the following kind:

underlying form	stat	'town'	state	'Pl.'	nagel	'nail'	nagele	'Pl'
umlaut	--		stäte		--		nägele	
lengthening	--		state		nagel		nagele	
rounding	--		--		ngel		--	

Various further rules combine to yield the phonetic output, which has several interesting features (including a predictable tone distinction in long vowels); I ignore these to simplify matters and consider only the representations brought about by the three above rules.

What happens now (e.g. Mackel 1905-7) is that the umlaut vowel in some dialects is replaced by  wherever it corresponds to an unumlauted form with phonetic . Thus stäte remains but ngele becomes ngele. The problem is that in this changed grammar stäte requires that umlaut should apply before rounding, whereas ngele requires that umlaut should apply after rounding. What is going on is that umlaut comes to apply to the singular phonetic form rather than to the base form. Formally, this corresponds to making the rules apply cyclically, in the original mutual order, first to the stem and then to the whole word:

underlying form	(stat)	((stat)e)	(nagel)	((nagel)e)
umlaut	--	--	--	--
lengthening	--	--	ngel	ngel
rounding	--	--	ngel	ngel
umlaut	--	stäte	--	ngele
lengthening	--	state	--	--
rounding	--	--	--	--

This example appears to indicate that cyclical application of rules is the unmarked, simpler situation and that it represents a special restriction for a rule not to apply cyclically. This is confirmation for the proposal of Halle and Chomsky that all rules are included in the cycle, non-cyclical application being insured by additional conditions placed on individual rules when this is needed.

In this chapter I have discussed mainly cases where either the form or the order of the rules, but not both, is affected, partly because such cases are most common in practice and partly because they throw the problems involved into clearer perspective and are thus ultimately more rewarding objects of study. I should stress here, however, that there is no reason whatsoever why imperfect learning could not involve change in the order and form of rules at the same time. It is even theoretically possible (though I know of no actual case) that imperfect learning should result in altogether radical

changes in the grammar. The possibilities of imperfect learning are constrained by the two conditions only that the changed grammar should be simpler than the original grammar (or, in the case of very small changes, as simple as the original grammar) and that it should generate nearly the same language. Usually, the joint effect of these two conditions is to limit imperfect learning to just the sort of changes in order or form of rules that I have devoted most of this chapter to. But there is no reason at all why much more drastic reformulations of the grammar should not take place by imperfect learning provided that the two conditions are met.

Also, there is no reason why imperfect learning should not involve underlying representations rather than rules. Change in underlying forms is in fact a rather common, but usually uninteresting form of imperfect learning. Let me illustrate it with a traditional example. The Latin change of intervocalic *s* to *r* resulted in changes such as honōs (Nom.) : honōris (Gen.). This rhotacism rule is part of Latin morphophonemics (iūs:iūris, genus:generis), but in certain words (masculine and feminine polysyllables) *r* has been generalized throughout the paradigm, so that we get e.g. Classical Latin honor:honoris. The change is simply that /honōs/ has been replaced by /honōr/ as the underlying representation.

In conclusion, instead of characterizing linguistic change by the single mechanism of innovation in relatively unconstrained form, it seems necessary to do it by two distinct mechanisms, innovation and imperfect learning, both of which are heavily constrained and which jointly define the possibilities of linguistic change in a much more satisfactory way than rule addition alone could do. Both innovation and imperfect learning can be given a simple and general characterization in terms of independently motivated notions of generative grammatical theory, a fact that provides further justification for the correctness of this theory.

## CHAPTER 3

### RECONSTRUCTION

#### 3.1 Relative Chronology

So far I have looked at phonological change mainly from the viewpoint of its bearing on linguistic theory rather than on problems of linguistic reconstruction. But the material presented in Chapters 1 and 2 does have corollaries which are relevant to this latter, equally traditional goal of the study of linguistic change. The present chapter is an attempt to outline some of these corollaries and to illustrate them with a discussion of Grassmann's Law, a case where generative grammatical theory leads to a historical account which differs from the traditionally accepted one in many respects.

The aspect of reconstruction on which I will focus in these remarks is the determination of the relative chronology of the innovations that a language has undergone. Traditionally, to approach questions of relative chronology is regarded as some kind of a supplementary refinement of the main task, which is conceived as the reconstruction of a proto-language and of a list of innovations that have converted it into its descendant languages. But in fact it is completely impossible to do historical linguistics without taking account of relative chronology, as the correct form of the innovations critically depends on the order in which they applied.

In Chapter 1, I described how phonological change was viewed in the linguistic theories of Paul, Saussure and Bloomfield, the principal spokesmen of three approaches to language within which much of the empirical work on historical linguistics in the twentieth century has been carried out. I suggested that in spite of great theoretical differences, the three place largely equivalent constraints on phonological change. These constraints amount to restricting phonological change to two forms: gradual drift along the phonetic parameters, and non-gradual replacement under certain highly

restricted phonetic conditions, e.g. assimilation, dissimilation, metathesis, etc. Of course the kind of change that I ascribed to imperfect learning in Chapter 2 has not gone unnoticed, at least in the more obvious cases. The trouble is rather that it has not been accounted for in theories of linguistic change and has not been taken seriously enough to be considered as a relevant factor when the relative chronology of prehistoric sound changes is determined. Therefore it is the case that in special investigations of relative chronology (e.g. Bremer, 1894; Goetze, 1923) and in historical descriptions which pay careful attention to relative chronology (e.g. Luick, 1921; Lejeune, 1947) the approach is based on the following implicit assumption, which is a direct consequence of taking sound change, in the taxonomic sense, to be the only form of phonological change:

The phonological changes that a language has undergone between stages A and B should, when applied in the order of their relative chronology to the phonetic forms of stage A, yield the phonetic forms of stage B.

In the special case where innovations lead to no restructuring, but are retained as synchronic phonological rules with unchanged underlying representations, this amounts to the following claim:

The synchronic order and form of rules reflects their historical order and form.

The material of Chapters 1 and 2 suggests that this special case, and hence the whole assumption, is false, mainly for the following three reasons.

1. Placement of innovations in non-final position. For example, the synchronic order of Lachmann's Law and the Indo-European voicing assimilation rule is the reverse of their relative chronology, because Lachmann's Law was entered in position before the voicing assimilation rule (Sect. 1.5).

2. Reordering. For example, umlaut in many German dialects has been moved so as to follow various historically much later rules. (Chapter 2).

3. Loss. For example, the modern English vowel shortening rule goes back to Old English, but had a quite different form there. (Ex. 4, Ch. 2).<sup>1</sup>

At first blush this might seem to make linguistic reconstruction a rather hopeless task. It is true, I think, that the possibilities of recovering past stages of a language are rather more limited than is commonly assumed. However, there are two kinds of language-internal considerations that give us a foothold in linguistic history: the constraints on innovation and imperfect learning on the one hand, and the fact of restructuring on the other. Together, these two determine the possibilities and the limits of internal and comparative reconstruction.

The most important constraints on innovations are inherent in the treatment of them as rules. A large part of Chapter 1 was devoted to providing justification for this idea of Halle's by showing that the independently motivated generative grammatical form of synchronic phonological rules conforms to the facts of sound change in many cases where the taxonomic conception of sound change does not. Beyond this, however, there appear to be additional constraints that have to do with the place at which rules can be adopted in the grammar. In section 1.6 I argued that these limitations have a psychological rather than a sociological basis. As one possibility I suggested that there might be an upper limit on the way in which the placement of new rules affects the application of the other rules to the representations of the language. The rather strong condition which I considered, namely that new rules cannot affect the application of the rules in the same component at all, i.e. they cannot introduce or remove representations that are analyzed by later rules in the same component at all, i.e. they cannot introduce or remove representations that are analyzed by later rules in the same component, may, like several of its predecessors, turn out to be false as more data accumulates. However, the correct condition undoubtedly lies along these same general lines.

The one thing that seems certain about imperfect learning is that its directionality is determined by simplicity. It seems to me that this is perfectly clear both in the case of loss and in reordering. What is debatable about the latter is merely which of several possible, to a large extent equivalent characterizations of the asymmetry of rule order is the right one, and the related question of how the greater simplicity of one type of order is to be formally reflected in linguistic theory. The characterization of the asymmetry

in terms of maximal applicability suggested in 2.3 seems to me to be rather well motivated, however, as does its proposed incorporation in linguistic theory in terms of the marked - unmarked distinction.

Another basis for linguistic reconstruction is given by restructuring. Thus, when a sound change leads to revision of the underlying representations, its chronology with respect to related sound changes is recoverable. For example, the Germanic sound change of voiced to voiceless stops (skabjan > skapjan) led to a corresponding restructuring in the underlying representations. Hence we know that this de-voicing must necessarily have taken place after another phase of Grimm's Law, in which voiceless stops became aspirated except after obstruents (skaptas > skap<sup>h</sup>tas). Otherwise the aspiration (and the later spirantization, e.g. skap<sup>h</sup>tas > skaftas) would have affected the p in skapjan and we should have had Germanic "skafjan" rather than skapjan.

If the restructuring is just partial and affects unproductive forms only, subsequent processes of imperfect learning may leave these unproductive forms intact as "relic forms". Thus, in Ex. 3 of Chapter 2, e's which historically but not synchronically were derived from a (e.g. net from \*nati) became phonemic and did not participate in the subsequent lowering to ä which affected the e's that were synchronically derived from a.

These are the kinds of considerations that typically must figure in linguistic reconstruction. In the following discussion of some innovations in Greek and Sanskrit I will try to illustrate the extent to which they, on the basis of explicit generative grammars, can help to put some order into historical problems of considerable complexity.

### 3.2 Grassmann's Law: Synchronic description

In both Greek and Sanskrit, Grassmann's Law (G.L.) is a phonological rule that eliminates aspiration before a following aspirate. E.G.:

Greek	thi-thē-mi > tithēmi,	the-throph-a > tetrapha
Sanskrit	dha-dhā-mi > dadhāmi	bho-bhudh-ye > bobudhye.

Although I will write aspirated stops in the usual manner as ph, bh etc., they are in fact single phonemes distinguished from unaspirated stops by the feature "tense-lax". Accordingly, G.L. is to be stated in the following form:

(G.L.' )



Both languages also have a Cluster Rule (C.R.) that eliminates aspiration immediately before obstruents:

Greek      gluph-s-ō → glupsō

Sanskrit labh-sya-te > labsyate (lapsyate by the later voicing assimilation rule).

(C.R.) [-vocalic] → [-tense] / \_ [+obstruent]

There are additional restrictions on these rules, to which I return directly. First note how the correct ordering

1. C.R.  
2. G.L.

accounts for the apparent "throwback" of aspiration to the initial stop in certain roots when the final stop of these roots is deaspirated, e.g.

Greek      trephō,    Fut. threpsō

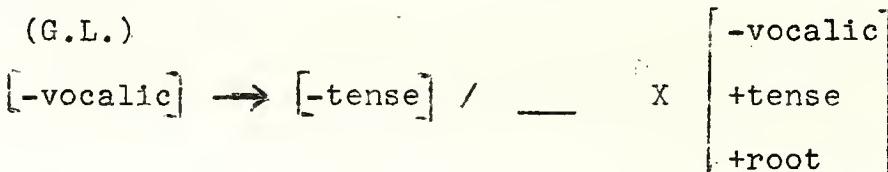
Sanskrit      dohmi,      2.p. dhoksi.<sup>2</sup>

The behavior of such roots is a consequence of the two rules and their ordering if we assume that their initial and final stops are both aspirated in the underlying representation. The correct phonetic forms are then derived as follows:

Underl. repr.	threphō	threphsō
C.R.	-	threpso
G.L.	trephtō	-

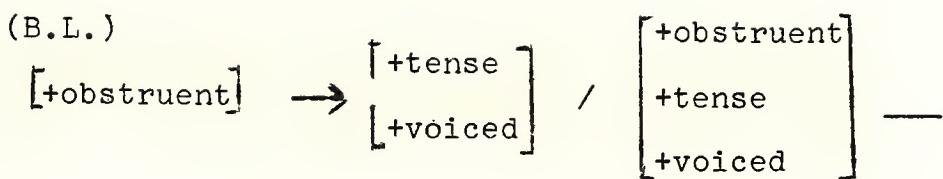
I now turn to a more detailed inspection of the form of C.R. and G.L. in Greek and Sanskrit, and the way they interact with other phonological rules in these languages.<sup>3</sup> Some quite striking similarities as well as many differences will emerge as the particulars of the picture in each language are filled in.

One important restriction in the application of G.L. in both Greek and Sanskrit is that the aspirate in an ending have a dissimilatory effect on an aspirate in a root or a prefix. Cf. Gk. Imper. phathi, phanēthi, straphēthi, Perf. Inf. pephantai, Aor. Pass. ephanthēn; Skt. vibhubhis, dhestha.<sup>4</sup> Hence G.L. must be restated in the following form for both languages:



The restriction that the second aspirate must be in the root interacts in an interesting way with the cluster rule to explain the fact that in Sanskrit as well as in Greek the initial segment of a double-aspirate root is not de-aspirated when the ending begins with an aspirate. Thus, in Sanskrit dhugh-dhve the cluster rule de-aspirated the root-final segment, yielding the correct phonetic form dhug-dhve. G.L. fails to apply to it since the remaining aspirate in the ending does not qualify as an environment.<sup>5</sup> Exactly the same situation prevails in Greek. The Aor. Pass. of the root threp is ethrepthēn, derived from underlying ethrephthēn by application of the cluster rule but not of G.L.<sup>6</sup>

In Sanskrit, the relation to G.L. of a well-known change which affects clusters of the form "voiced aspirate + stop" to yield clusters of the form "voiced unaspirated stop + voiced aspirated stop" is of importance. Examples: rundh-tas → runddhas, rundh-thas → runddhas. This change consists of two steps. The first is a progressive assimilation of voicing and aspiration. We will call this step Bartholomae's Law (B.L.), which is the name usually applied to the whole process rather than just the assimilation.



This is followed by the second step, deaspiration of the first aspirate of the resulting biaspirate cluster by the already discussed C.R.

Sample derivations:

	rundh-tas	rundh-thas
B.L.	rundh-dhas	rundh-dhas
C.R.	rund-dhas	rund-dhas

Diaspirate roots undergo both G.L. and B.L., e.g. dhugh-tas → dugdhas. The mutual order of G.L. and B.L. is irrelevant but both must precede C.R.:

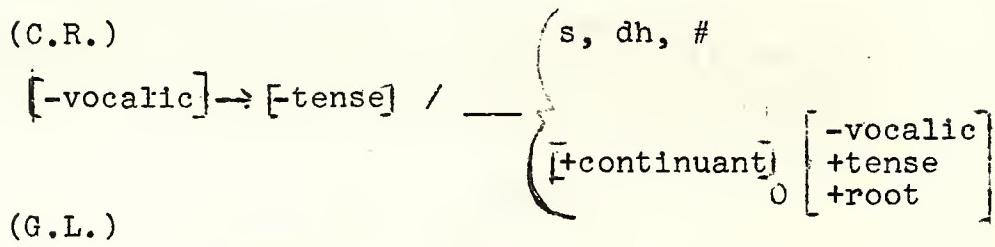
	dhugh-tas	dhugh-thas
G.L.	dugh-tas	dugh-thas
B.L.	dugh-dhas	dugh-dhas
C.R.	dug-dhas	dug-dhas

But this conflicts squarely with our former observation that G.L. has to be applied after C.R. in a derivation such as

	dhugh-dhve
C.R.	dhug-dhve
G.L.	-

Even cyclical application of the rules as they stand will yield incorrect results. If C.R. applies before G.L. and B.L., dhugh-tas will incorrectly become "dhugtas" ("dhuktas" by the later voicing assimilation rule.)<sup>7</sup> If, on the other hand, G.L. and B.L. apply first, dhugh-dhve yields the incorrect "dugdhve". The simplest set of rules seems to emerge if C.R. is restricted so as to apply before s, dh and word boundary, but not before t, th, and sonorants.<sup>8</sup> The rules apply cyclically, with the

revised C.R. applying first.<sup>9</sup> This enables us to combine G.L. and C.R. in a single rule with two subparts (in marked order) and thus to relate the two processes of deaspiration. We get the following Sanskrit system:



### Derivations:

dhogh-si      dhogh-ti      dhugh-dhye

### First cycle:

(C.R.)	dhog-si	-	dhug-dhve
(G.L.)	-	dogh-ti	-
(B.L.)	-	dogh-dhi	-

### Second cycle:

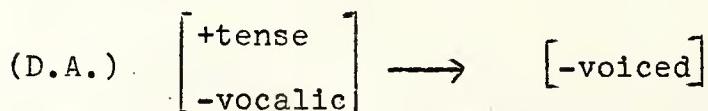
(C.R.) - dog-dhi -

In Sanskrit, voicing is distinctive for aspirates, and the deaspirated stops retain the voicing of the underlying form:

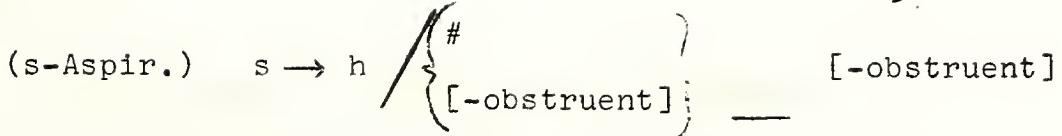
cha-chand-a → cachanda

dha-dhar-a → dadhara

But in Greek, all aspirates are predictably voiceless. G.L. is here preceded by an Aspirate Devoicing rule (D.A.):



Another Greek rule which is of importance in connection with G.L. is the rule which turns s into h in sonorant environments:



The root sek̥, which retains the s intact in the aorist e-skh-on shows it as h in the future hek-sō. The h that results from this rule drops by G.L. in the same environment as the other aspirates. Thus, in the present tense of the same verb, the underlying sek̥h-ō becomes hek̥h-ō by s-Aspir. and then ekhō by G.L. Note that we do not have to add a new environment to G.L. or change it in any other way in order to accomodate this new case of deaspiration, since h in Jakobson's feature system is a tense glide. Application of G.L. will turn it into the corresponding lax glide - the "smooth breathing" of the ancient Greek grammarians.

In Greek as in Sanskrit C.R. and G.L. can be combined into a single rule with two subparts in marked order. D.A. and s-Aspir. must both precede this rule, this being in each case an unmarked order. D.A. and s-Aspir. are mutually unordered. No cyclic application is needed.

Derivations:

The-ThroPh-a	e-ThrePh-Thēn	ThrePh-s-ō	seKh-ō	seKh-s-ō
(D.A.)				
the-throph-a	e-threph-thēn	threph-s-ō	sek̥h-ō	sek̥h-s-ō
(S-A).	-	-	-	hek̥h-ō    hekh-s-ō
(C.R.)	-	e-threp-thēn	threp-s-ō	-    hek-s-ō
(G.L.)	-	-	-	ekh-ō    -
te-troph-a	-	-	-	

(By a capital letter I denote a segment not specified for voicing).

In summary, Greek and Sanskrit both have G.L. in identical form. In both it is combined with a similar form of C.R., which, however, is more restricted in Sanskrit. The other main differences are:

1. Sanskrit but not Greek has B.L.
2. The rules operate cyclically in Sanskrit but not in Greek.
3. In Greek, G.L. is preceded by two rules not found in Sanskrit, namely D.A. and s-Aspir.

### 3.3 Grassmann's Law: Internal Reconstruction

Certain acts of internal reconstruction can be undertaken purely on the basis of the synchronic descriptions developed in the previous section.

First, we can establish the fact that C.R. (in some form) must have arisen earlier than G.L. in both Greek and Indic. To see why this is so, note first that the order

1. C.R.

2. G.L.

is marked order by the condition proposed in Section 2.3, which therefore could not have arisen by reordering. Secondly, the condition proposed in section 1.6 excludes the possibility that C.R. should have been entered in the grammar later in time than G.L. but in position before it. Furthermore, if G.L. had applied before C.R. it would have deaspirated the initial aspirate of diaspirate roots in all environments. But then there would have been no basis for subsequent reintroduction of the initial aspirate. Thus the relative chronology can be supported in both of the ways mentioned in the previous section: by an argument from the limitations on innovation and imperfect learning, and by an argument from restructuring.

From this we would go on, with much less certainty, to assume that cyclic application of the rules in Sanskrit was preceded by a stage in which they applied non-cyclically. The subsequent change to cyclic application, by the form of simplification discussed in 2.5, would then parallel very closely the German example cited there.

Internal reconstruction can be pushed still further back in Sanskrit if other facts are taken into consideration. It has long been known (Wackernagel, 1896, 126) that double-aspirate roots in Sanskrit occasionally lose both aspirates before endings beginning with *s*, in direct conflict with the rules of Section 3.2. This happens mainly in certain rare or morphologically opaque words, e.g. *grtsas* 'dexterous' (historically but hardly synchronically related to *ghrdh-* 'be eager'),

the rare 3.Pl. bapsati (bha-bhs-atı, from the root bhas 'chew') etc. These anomalous forms, which are most frequent in the oldest texts (Rigvedic), are evidently relic forms reflecting an earlier version of Indic phonology.

Our problem is to find a grammar that meets the following two conditions: 1) it must generate the relic forms (grtsas etc.) from their etymological representations (ghrdhsas etc.); 2) It must be a possible ancestor of Sanskrit, i.e. Sanskrit must be derivable from it through bona fide forms of linguistic change.

The traditional explanation for relic forms like grtsas is that B.L. and C.R. (it will be recalled that these are combined into a single step in the usual formulation) originally preceded Grassmann's Law, with derivations such as the following:

	<u>ghrdhsas</u>	<u>dhoghsı</u>
(B.L.)	<u>ghrdhzhı</u>	<u>dhoghzı</u>
(C.R.)	<u>ghrdzhas</u>	<u>dhogzı</u>
(G.L.)	<u>grdzhas</u>	<u>dogzı</u>

Subsequently a sound change deaspirated and devoiced the clusters dzh, gzh, bzh to ts, ks, ps, yielding grtsas, doksi. By analogy, doksi then became dhoksi. (Thumb-Hauschild, 1958, 297; see also the references to Brugmann and Wackernagel cited there).

The simple but fatal flaw in this solution is that G.L. would not apply to the third line in the above derivation to yield the bottom line, for the same reason that it does not apply to dhugdhve to give "dugdhve" or to vibhubhis to yield "vibubhis", namely because of the restriction, well motivated both in Greek and Indic, that the aspirate in the environment must be in the root, as pointed out in 3.2. The solution is thus based on an incorrect formulation of the facts and must hence be rejected.<sup>10</sup>



1) clusters like dhzh became deaspirated and voiceless, e.g. ts; 2) an underlying initial aspirate was phonetically restored wherever a following aspirate was lost due to change 1). I propose that both these changes were a single change resulting from the generalization of (C.R.') to (C.R.).

This analysis raises an interesting question regarding opaque, isolated forms of the type gapsas in which the initial aspiration was neutralized because there was no morphophonemic support for either aspiration or non-aspiration in the underlying representation. Before the simplification of C.R., a hypothetical opaque word like gapsas could, when the root was not otherwise found in the language, be given two underlying representations, both of which were compatible with its phonetic form and equally simple. The analysis gabh-sas and the analysis ghabh-sas would at this point have resulted in the same phonetic form gabhzas. But as C.R. was generalized, the first analysis yielded phonetic gapsas and the second ghapsas. Is any trace of such vacillation in opaque forms preserved?

There are only a few cases of such isolated words, but in one of them we do in fact find vacillation of this kind. The word drāksā 'vine, grape' has a doublet dhrāksā (unexplained, so far as I know). I suggest the above as a way of accounting for the two forms.

### 3.4 Grassmann's Law: Comparative Reconstruction

Up to now I have viewed Greek and Sanskrit independently of each other, trying to push as far back as possible into the history of each through internal evidence alone. In this section the question will be taken up of what historical connection, if any, there is between G.L. in Greek and G.L. in Sanskrit.<sup>11</sup> It was raised already by Grassmann himself in 1863. His conclusion, reached with obvious reluctance, that the Greek and Indic dissimilations were historically unrelated has remained unquestioned ever since.<sup>12</sup> Neither neogrammarians nor structuralists have seen any need to re-examine the argument that underlies this dismal conclusion

in the light of subsequent insights into linguistic structure and linguistic change. The classical line of argument was concisely summarized by Bloomfield as follows:

"A further step in the reconstruction of the historical events' proceeds from the facts that the loss of aspiration results in Sanskrit in [b, d, g], but in Greek in [p, t, k]. This implies that the Primitive Indo-European [bh, dh, gh] had already become unvoiced [ph, th, kh] in pre-Greek when the loss of aspiration took place. Since this unvoicing does not occur in Indo-Iranian, we conclude that the de-aspiration in pre-Greek and the de-aspiration in pre-Indo-Iranian took place independently." (1933, 351)

The argument has two steps:

- 1) The deaspiration took place after the Greek aspirated became unvoiced since it yields voiceless stops in Greek.
- 2) Since the deaspiration took place after the Greek aspirates became unvoiced it took place independently in the two languages.

The same argument, with both steps, could be repeated with respect to G.L. and s-Aspiration:

- 1) The deaspiration took place after Greek s became h since it applies to this h.
- 2) Hence it took place independently in the two languages.

I want to argue in the following that both steps of these arguments are false, and that, furthermore, the second step cannot even be defended within Bloomfield's linguistic theory. Finally, I want to propose an account which not only relates the Greek and Indic dissimilations but also sheds light on certain apparently anomalous phonological developments in Greek.

Taking the second step of the above argument first, let us inquire after the rationale that underlies it. By what general considerations could one support the contention that the specifically Greek sound changes (D.A. and s-Aspir.) preclude Greek

and Indic from subsequently borrowing a sound change (G.L.) from each other? There is no hint of this in Bloomfield, and other statements on the subject, though wordier, are no more informative.

What immediately comes to mind is some strong form of the Stammbaum theory, say the contention that dialects which differ in some respect cannot undergo joint innovations, i.e. that isoglosses cannot intersect.<sup>13</sup> From this, the second step follows directly. It is entirely possible that this is what underlay Grassmann's reasoning in 1863. Yet very few linguists of the twentieth century would subscribe to this view; certainly not Bloomfield:

"The comparative method presupposes clear-cut splitting off of successive branches, but the inconsistent partial similarities show us that later changes may spread over the isoglosses left by earlier changes; that resemblance between neighboring languages may be due to the disappearance of intermediate dialects (wave-theory); and that languages already in some respects differentiated may make like changes." (1933, 318)

The discussion preceding this passage appears to indicate that Bloomfield considered only the spreading of innovations over closely related dialects, and not over language boundaries. Such a restriction is quite unmotivated and contradicted by a host of facts.<sup>14</sup> But whether or not Bloomfield had some such restriction in mind is not relevant, for it can in any case not have played a role in the case of G.L., since the two specifically Greek changes cannot obviously have split Greek and Indo-Iranian into two separate languages.

But there is another possible, quite distinct reason why the Greek devoicing of aspirates should have been thought to preclude a subsequent common Greek-Indo-Iranian deaspiration, even when the falsity of Stammbaum theory is admitted. Sounds have commonly been regarded as unanalyzable entities rather than as complexes of distinctive features. But then there is nothing at all in common between G.L. in Greek and G.L. in Sanskrit. The only thing that can be said is that Greek

changes ph, th, kh into p, t, k respectively, whereas Sanskrit changes bh, dh, gh into b, d, g respectively. Under this view, these are entirely separate events which could not in principle be related to each other. We may say, then, that in general a corollary of denying the feature structure of sounds is that a sound change can be borrowed from one dialect into another only if the domain in which it applies functionally is the same in both. This, in effect, trivializes the wave theory into an "isogloss theory" by restricting it, for practical purposes, to the spread of lexical items and the like.

Again, this argument may have played a role in the early days, but clearly not for Bloomfield, to whose conception of phonology distinctive features were crucial. On p. 353 of Language he specifically remarks that features ("habits of articulation... common to several phonemes") are themselves subject to change. And obviously the whole discussion of G.L. on pp. 349-51 shows that he considers it to be the same sound change in both languages.

There was thus no reason at all for Bloomfield to deny that G.L. could have "spread across the isogloss" created by D.A. and s-Aspir. in Greek. The fact that he did must be simply due to the venerable status of G.L. as a paradigm case of independent innovation. Götze (1923) considered it a "Musterbeispiel" of an isogloss that could be shown by the methods of linguistic reconstruction to be spurious. Traditional conclusions remain standing long after their foundations have been exploded.

Let me now return to the first step in the argument, the one purporting to show that G.L. could not have arisen in Greek before the aspirates were unvoiced and s became h. It furnishes a good illustration of an application of the principle that I claimed in 3.1 to underlie traditional reconstruction of relative chronology, namely that the phonetic representations of the descendant language should be derivable from the phonetic representations of the ancestor language by application of the sound changes in the order of their relative chronology. I would now like to show how in this particular case there is good reason for altogether different conclusions.

The two Greek innovations under discussion, D.A. and s-Aspir., are in unmarked order with respect to G.L. It is accordingly a possibility that G.L. originally preceded them and is historically earlier. If so, G.L. may well have been an innovation that took place before the Greek-Indo-Iranian dialect area of IE diverged. At a certain stage of Pre-Greek we then had C.R. and G.L. followed by the two innovations D.A. and s-Aspir.:

- (1a) C.R.
- (1b) G.L.
- (2) D.A.
- (3) s-Aspir.

At this point, the deaspiration by Grassmann's Law still yielded voiced stops which alternated with voiceless aspirates. In place of classical Greek trep̄ō:threp̄so we still had drep̄ō:threp̄so. This alternation between voiced unaspirated stops and voiceless aspirated stops motivated the retention of underlying voiced aspirates and D.A. as a morphophonemic rule. Also, at this point the aspirate derived from s did not dissimilate: we had hek̄ō in place of classical Greek ek̄ō.

D.A. and s-Aspir. stood in a marked order with respect to G.L. and both were hence susceptible to inversion with it. The inversion of D.A. and G.L. had the effect of replacing the two-feature alternation in drep̄ō:threp̄so by the familiar one-feature alternation of Classical Greek, and also permitted further simplification in the morpheme structure rules. The inversion of s-Aspir. and G.L. had the effect of making the aspirate h from s undergo the same dissimilation by G.L. as the other aspirates. Together, these two order changes give us the attested situation of classical Greek.

Thus we do not, after all, have to admit the extraordinary coincidence of a rule as striking as G.L. arising independently in two IE dialects as closely related as Greek and Indic. In fact, we are now even in the position

of having to choose between two entirely distinct ways in which the Greek and Indic dissimilations might be historically connected. As always in choosing between competing theories, the problem is to find crucial cases that will discriminate between them. A critical difference between the two accounts is that only one of them (the one which supposes G.L. to antedate D.A.) assumes the former existence of forms like drehō, replaced by the classical Greek trephō. Would relic forms have survived in Greek which have escaped this replacement?

Let us look more closely at the stage of Pre-Greek with G.L. and D.A. in marked order which this account postulates. The morphophonemic distinction between the IE root types ghedh and gedh would at this point still have been retained to the extent that these roots occurred before an obstruent or word boundary that caused loss of their second aspirate by C.R. Thus, [drehō] would have been represented as /dhrebhō/ with an initial underlying aspirate, and [gluphō] as /glubhō/ with an initial underlying non-aspirate, because of the distinction in the future forms [threpsō], [glupsō]. This would have been the case quite generally in root nouns, verbs with sigmatic tenses etc. But there were also roots that never happened to occur in any such pre-obstruent environment. In these, the distinction between the types bhedh and bedh must have been irrevocably lost for want of morphophonemic support. In such opaque forms the operation of G.L. rendered the initial stop ambiguous or neutralized as to underlying aspiration. Whether given the morphophonemic representation /gedh/ or /ghedh/ they obtained the phonetic form [geth] by G.L. and D.A.

Let us reconstruct the further development of this hypothetical stage of Pre-Greek. When the devoicing rule D.A. came to apply before G.L. the outcome in derived forms, we recall, was that the morphophonemic difference in the underlying form obtained a phonetic reflex: [drehō] (cf. [threpsō]) became [threphō] while [gluphō] (cf. [glupsō]) stayed unchanged. In opaque forms, however, the original distinction between initial aspirate and non-aspirate had been lost and could not re-emerge. We should instead expect the underlying ambiguity

to be reflected in phonetic vacillation with no systematic relation to etymology.

This provides a crucial test for determining the relative chronology of G.L. and D.A. If opaque forms retain the etymological distinction between the root types gedh and ghedh, D.A. was earlier than G.L., if not, it was later. As test cases, then, we require words which are synchronically isolated in Greek and which by means of clear cognates in Germanic, Italic or some other IE dialect that preserves reflexes distinctive aspiration can be shown to have had IE roots of the type gedh or ghedh.

The majority of the words that qualify as test cases are of the type ghedh. It appears that the etymological distinction between initial aspirate and initial non-aspirate is indeed obscured. In some cases we get an initial voiceless stop:

Dor. pákhús 'arm'; OIcel. bógr; Skt. báhús.

In others we get a voiced stop:

bóthros 'pit'; L. fodíō 'dig'; Engl. bed.

brekhmós 'front of head'; OE brægen 'brain'.

In a third set of cognates, both voiced and voiceless reflexes are actually attested in Greek itself:

agathós 'good'; Engl. good.  
Variant akathós in Hesychius.

buthós 'depth, abyss', L. fundō;  
same root with p- in

puthmén 'ground, base, foundation'.

Roots of the gedh type were rarer in IE than those of the ghedh type. Consequently there are very few test cases of this type available.<sup>15</sup> I know of only two completely clear cases: dolikhós 'long' (cf. Goth. tulgus) and gómpbos 'nail, peg', gomphíos 'molar tooth'<sup>16</sup> (cf. Engl. comb). In the latter case, a doublet kómpos 'molar tooth' (with the common lenition after nasals) is in fact preserved in Hesychius.<sup>16</sup>

The evidence is thus not plentiful but nevertheless suggestive. We do not find (nor would we expect to find) doublets preserved in every case. But we do find - and this is crucial - that the etymological distinction between the root types gedh and ghedh is not systematically preserved in Greek where it had no morphophonemic support from the cluster rule. If this is a fact - and I do not see how it could be refuted except by giving better etymologies for the crucial examples<sup>17</sup> - it clinches the early operation of G.L. in Greek, and makes it likely that G.L. was a rule shared by the Greek-Indo-Iranian dialect area of IE before it split up into separate dialects.

FOOTNOTESChapter 1

- 1) Saussure (1959, 83-7), Jakobson (1929; 1931), Halle (1962).
- 2) Saussure (1959, 83-5), Twaddell (1938), Penzl (1949). Needless to say, my discussion of umlaut here is extremely schematic and I omit, for example, consideration of secondary umlaut. This will not affect the point being made.
- 3) I cite this example here merely to illustrate the mechanics of restructuring, ignoring at this point the serious questions of relative chronology that must be raised with regard to the different steps of Grimm's Law. To some of these I return in Chapter 3. The facts are clearly presented in Streitberg (1963); for illuminating comments, see Halle (1961; 1962).
- 4) The exact difference is that secondary suffixes are preceded by word boundaries, i.e. act as enclitics. This fact has many striking reflexes in English phonology: for example, /ng/ retains its word-final form [ŋ] before agentive -r. (hence singer [siŋ'r] but finger [fiŋ'gr]). See Chomsky and Halle (forthcoming).
- 5) This whole passage of Bloomfield's is rejected by C. Hockett, a well-known American linguist (1965, fn. 20). His theory of sound change is essentially an impoverished version of Paul's and Bloomfield's, with articulatory drift as the only form of sound change. Hockett perhaps intended to meet the obvious objection that metathesis,

assimilation, dissimilation etc. cannot conceivably be due to articulatory drift by his bizarre though unsupported remark that these are "allied more closely to analogy than to borrowing or to sound change" (1958, 390). To clarify this, it should be remarked that Hockett distinguished three mutually exclusive forms of linguistic change: sound change, analogy, and borrowing; rather as if one were to partition the class of animals into mammals, insects and pets. One of the consequences of Hockett's view would be that sound change cannot be borrowed; another would be that syntactic and semantic change either do not exist or are always due to analogy or borrowing. It is not clear to me whether Hockett actually would accept these absurd consequences. For a commentary on Hockett's theory of sound change, see Postal (forthcoming).

- 6) In the Rumanian case this has in fact been attempted many times, e.g. by P. Naert (1941), who suggested the following tortuous but undeniably gradual developments:

$$\begin{aligned} kt &> xt &gt; \overset{x}{wt} &> \overset{w}{t} &> \overset{y}{t} &> pt \\ gn &> \overset{g}{n} &> \overset{\circ}{wn} &> wn &> \overset{\circ}{n} &> mn \end{aligned}$$

The fact that Latin orthographic and phonemic gn was pronounced [b̥n] not only completely disrupts the parallelism of the two changes in Naert's account but also makes the postulated spirantization extremely implausible from the phonetic point of view. To preserve the parallelism one would have to suppose that [j̥n] first somehow reverted to [gn], thus piling one ad hoc assumption on another. It baffles me why Naert's account has been taken seriously (Rothe, 1957; Hoenigswald, 1960, 55).

- 7) Halle (1962 a), Bever (forthcoming).
- 8) Rohlfs (1949, pp. 48, 136, 151, 185, 213).
- 9) A similar point is made and supported with an example from Swiss German by Moulton (1960).
- 10) C. Lachmann, In Lucretium Commentarius, I.805: "...participia passiva ea quarum in praesenti

consonans est aut liquida aut s semivocalis,  
 quantitatatem praesentis secuntur... contra  
 ubi in praesenti media est, participia  
 producuntur."

- 11) I must still meet the possible objection that an IE \*set<sup>S</sup>tłom became \*setłom somehow by analogy to forms with sed- independently in all the Indo-European languages, somewhat like Sanskrit dehi has an analogical doublet dadd<sup>n</sup>i in Vedic. In anticipation of the material of Chapter 2, I point out that this would normally be a reordering of the dental assimilation rule to precede (32) and (33). This objection can be met easily enough as the same treatment is found in other words even when languages have otherwise lost the root, thus depriving the supposed analogy of a model.
- 12) I am excluding here the rules that assign the fine phonetic detail in terms of n-ary features; clearly these will apply to whatever segments a phonological innovation introduces. I assume that they form an entirely separate part of the phonology.

## Chapter 2

- 1) Thus, unumlauted ä is retained in part in the Kanton of Schaffhausen (Wanner, 1941). Cf. the similar standard German situation, where ei 'egg' contains a phonetic back vowel which never becomes fronted by umlaut.
- 2) E.g. the dialect of Kesswil, Oberthurgau (Enderlin, 1913).
- 3) A. Bachmann (ed.), Beiträge zur schweizer-deutschen Grammatik, Volumes 1, 2, 3, 5, 7, 8, 9, 10, 13, 14, 16, 20.
- 4) Weinreich (1963), Sapir (1915).
- 5) For example, by Klíma (1965). Alternatively, one could say that the rules become optional and are, then, simply by chance never applied. This in no way helps to meet my criticism.

- 6) My understanding of the rules of palatalization in the various Slavic languages is based on Halle and Lightner (forthcoming). For the data, see e.g. Vaillant (1950).
- 7) Chomsky (1965, Ch. 3, fn. 6); cf. also Zwicky (1965), McCawley (1965).
- 8) In most dialects, other than these dialects in the Northern fringe of Switzerland, two extensions of the lowering complete the picture: first, o is lowered to ɔ sporadically before grave consonants as well; second, o is regularly lowered to ɔ in past participles regardless of what consonant follows.
- 9) A very clear statement that it is individual words and not rules that figure in analogy is the following (Wheeler, 1885, 5): "Dass das Verbum ein Princip, eine Abstraction, durch den Einfluss des Nomens geradezu adoptieren sollte, ist undenkbar. Einzelne concrete lautliche Erscheinungen bewirken aber einen solchen Einfluss, der um so grösser sein kann, je zahlreicher die eine solche Erscheinung aufweisenden Fälle sind. Das Princip siegt nur nachträglich durch die dasselbe illustrierenden concreten Fälle, aber auch dann nicht nothwendigerweise oder immer in unveränderter Gestalt." This is in striking contrast to pre-neogrammarian usage, where analogy simply meant "rule", analogical change being called "false analogy".
- 10) Steps in this direction have in fact been taken by Kuryłowicz e.g. (1964).

### Chapter 3

- 1) The possibility of loss has also relevance to the reconstruction of relative chronology. For example, it is by no means necessary that Verner's Law, which made spirants voiced in a certain environment, should have applied after the Germanic spirants f θ x arose from IE p t k which had become aspirated. It is perfectly possible that Verner's Law applied before Germanic f θ x arose to the sole inherited spirant s and was subsequently generalized to apply to the new spirants when they developed.

- 2) In citing Sanskrit forms, I omit certain automatic phonetic details, writing e.g. dhoksi for dhokṣi, grtsas for grtsah. Except in etymological discussion, I also omit indicating accent in Greek and in Sanskrit.
- 3) For detailed generative phonological treatments of Grassmann's Law in Sanskrit and Greek, see Zwicky (1965) and Langendoen (1965). Cf. also Lightner (1966), for interesting comments on Saussure's treatment of Grassmann's Law.
- 4) We get phonological doublets in the case of ambiguous morphological structure. An original \*dhedhmos 'custom, law', from the root dhe 'set' was analyzable as either dhe-dhmos with root plus suffix -dhmos, or as dhe-dh-mos with reduplication, zero grade of the root, and suffix -mos. The former analysis underlies Lac., Arc., Locr. thethmos, and also Ion.-Att. thesmos; the latter gives Dor. tethmos. See Frisk (1960) s.v. The usual cock-eyed explanation is that the dissimilation of aspirates was offset by sporadic contrary assimilation of aspirates (Schwyzer, 1939, 262). It should be pointed out that there are a few exceptions to the restriction in Greek: notably the Aor. Passives etethēn, etuthēn.
- 5) The 2.Sg. Imperative ending -dhi poses a problem. Instead of the expected dhugdhi we get dugdhi. Note that the correct result would be obtained if the Imp. ending were treated as an enclitic rather than as an ending, as preceded by a word boundary. This is an ad hoc solution for Sanskrit, unless independent motivation can be found. It is nevertheless interesting in view of the Slavic parallel (Jakobson 1932, 355-6).
- 6) The spelling ethrephthen does not indicate a cluster of two aspirates. This is the ordinary Greek convention for writing a cluster of the form "unaspirated stop + aspirates stop".
- 7) Interestingly, Vedic has such forms (Whitney 1889, 55).

- 8) This modifies somewhat the solution proposed by Zwicky (1965), who places an analogous restriction on G.L. I leave open the question how the environment in which C.R. applies is to be characterized in terms of distinctive features. Because of the severe restrictions on what segments can be endings in Sanskrit, several alternatives are possible. Zwicky proposes the class of tense segments, with s thus regarded as being tense, and th as a cluster t + h, a step for which there is some independent motivation. An alternative would be to say that C.R. applies before segments that disagree in the features voiced and continuant. Thus, it applies before voiced non-continuants (dh) and voiceless continuants (s) but not before voiceless noncontinuants (t, th) and voiced continuants (vowels, glides). Actually, both alternatives are pretty ad hoc.
- 9) If Sanskrit has a special cycle for the stem without the ending, as Zwicky suggests, G.L. must be restricted to word boundary. This changes none of the derivations given, and merely involves one extra vacuous pass through the cycle.
- 10) I am challenging two traditional claims: one, that application of G.L. to such clusters like dzh suffices to explain relic forms like grtsas; this I have done above; the other, that the existence of such clusters can be established by internal and comparative reconstruction. This has been claimed by Wackernagel (1896, 239), one of his motivations being the false argument from G.L. Another concerns the development of underlying ja-ghs-ta to jagdha. He supposes that ja-ghs-ta first became jagzdha by B.L. and C.R. and that the sibilant subsequently dropped between stops by the well-known rule. In the first place, this would be, if correct, an argument for an optional s in the environment of B.L. rather than for the clusters dzh etc. which are at issue. In the second place, it is false. We may just as well suppose that ja-ghs-ta first lost the sibilant and the resulting ja-gh-ta subsequently became jagdha by B.L. and C.R.

The remaining arguments are comparative in nature and concern Pāli and Iranian forms.

They indicate the former existence of voiced clusters but are, so far as I can see, neutral as to whether they were of the form dzh, as assumed below, or dzh.

- 11) Another question, which has occasioned much more interest and controversy, will not be touched on here, although it would properly belong in this section if this account were to be complete. This is the question whether Greek formerly had B.L. and a more restricted form of C.R. There is a voluminous literature on this, but I know of no conclusive argument either way.
- 12) The first, to my knowledge, to challenge the traditional doctrine on G.L. was G.H. Matthews, who proposed essentially the first solution I discuss below.
- 13) The history of the notions 'wave theory' and 'stammbaum theory' is a confused one. The point of Schmidt (1872) was twofold: that isoglosses could intersect; and that isoglosses between the IE languages in particular did intersect. Both these claims came to be called 'wave theory' and their denial 'stammbaum theory'. In many of the early discussions around Schmidt's monograph the attention centered primarily on the specific IE implications. Thus, Leskien (1876) all but ignores the general question and seeks to criticize 'wave theory' by citing evidence of a historical kind that IE tribes which migrated became isolated. Later, as interest in IE dialectology subsided, the general question of intersecting isoglosses came to the fore. But by a curious terminological shift, stammbaum theory also came to denote not, as originally, the denial of wave theory, but instead the much weaker (and totally undisputed) assertion that dialects could lose contact and not undergo any joint innovations. This is evidently what is meant when it is said that wave theory and stammbaum theory are 'compatible' or 'reconcilable'.
- 14) Boas (1911), Jakobson (1931), Sommerfelt (1963), Emeneau (1956). The very languages we are dealing with provide examples in support of

this view. The Greek-Indo-Iranian change of syllabic nasals to a - a change so unusual that the possibility of independent development in each of the languages is highly unlikely - took place after s-Aspir. in Greek. Thus \*dnsus → Gk. dasús (Lat. densus); \*nsis → Gk. ásis (Lat. ensis, Skt. asis), cf. Thieme (1958). As these rules are in marked order and the change of syllabic nasals to a led to partial restructuring, we can be certain of their relative chronology.

- 15) Because of the heavy restrictions on Indo-European root structure, G.L. has some interesting structural consequences in both Greek and Sanskrit, which have been studied by Hoenigswald (1964) and Langendoen(1965).
- 16) The entry is: kóbous; odóntas gomphious. There are some much less certain examples, e.g. gniphón:kniphón 'miser, skinflint'. The connection with Lit. gnybu, German kneifen 'pinch' (IE \*gneibh-) seems reasonable and would account for the vacillation in the initial stop. But it may as well be due to the expressive character of the word.
- 17) The etymologies given are of course not new. Most of them were given by Grassmann in his original paper of 1863. They are not even rejected by the very cautious Frisk (1960), in spite of the phonological difficulties. Similarly, Schwyzer (1939, 262) notes that bóthros and fodiō "cannot be torn apart" and suggests that in this one word G.L. antedated the devoicing of aspirates. There are at least ten articles with more or less fanciful attempts to provide explanations for agathós and good, including two particularly entertaining ones by Matteo Bartoli. Needless to say, a claim that these words are not Greek but loanwords from some unknown Indo-European language, or from a lost dialect of Pre-Greek (perhaps "Pelasgian") would be completely ad hoc.

## BIBLIOGRAPHY

- Allen, W. S., "Relationship in Comparative Linguistics", Transactions of the Philological Society (1953).
- Benveniste, E., "Sur la phonétique et la syntaxe de l'armenian", Bulletin de la Societe de linguistique de Paris, Vol. 51 (1959).
- Bever, T. G., Theoretical Implications of Bloomfield's Manomini Studies. M.I.T. Dissertation (forthcoming).
- Birkeland, H., Growth and Structure of the Egyptian Arabic Dialect. Avhandlinger utgitt av Det Norske Videnskapsakademi i Oslo: II. Hist.-Fil. Klasse, No. 2 (1952).
- Bloomfield, L., Language. New York: Holt (1933).
- Boas, F., "Introduction", Handbook of American Indian Languages, Vol. I (1911). Page references are to the reprint, ed. C.I.J.M. Stuart (1963).
- Bremer, O. "Relative Sprachchronologie", Indogermanische Forschungen 4.8-31 (1894).
- Chomsky, N., "Current Issues in Linguistic Theory", in Fodor and Katz, The Structure of Language. Inglewood Cliff, N.J. (1964).
- Chomsky, N., Aspects of the theory of syntax, Cambridge, Mass. M.I.T. Press (1965).
- van Coetsem, F., Das System der starken Verba und die Periodisierung im älteren Germanischen. Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen. Afd. Letterkunde, Nieuwe reeks, Deel 19, No. 1 (1956).
- Cohen, M., "Categories de mots et phonologie", Travaux du Cercle Linguistique de Prague 8.36 ff. (1939).

- Djaukian, G.B., "K voprosu o proisxoždenii konsenantizma armjanskix dialektov", Voprosy jazykoznanija (1960), No. 6, 39-49.
- Emeneau, M.B., "India as a linguistic area", Language. 32.3-16 (1956).
- Enderlin, F., Die Mundart von Kesswil. Vol. 5 of A. Bachmann (ed.), Beiträge zur Schweizerdeutsch Grammatik (1913).
- Fodor, J. and J. Katz, The Structure of Language. Englewood Cliffs, N.J.: Prentice-Hall (1964).
- Foley, J., Spanish Morphology. Unpublished dissertation, M.I.T. (1965).
- Frisk, Hj., Griechisches etymologisches Wörterbuch I. Heidelberg: Wirter (1960).
- Garde, P., Review of American Contributions to the Fifth International Congress of Slavists, Word 21.141-7 (1965)
- Garibian, A.S., "Ob armjanskom konsonantizme", Voprosy jazykoznanija (1959), No. 5, 81-90.
- Goetze, A., "Relative Chronologie von Lauterschönungen im Italischen", Indogermanische Forschungen 41.78-194 (1923).
- Halle, M., "On the role of simplicity in linguistic description", in R. Jakobson (ed.), Structure of Language and its Mathematical Aspects, Proceedings of the XIIth Symposium in Applied Mathematics, pp. 89-94. Providence, R.I.: American Mathematical Society (1961).
- \_\_\_\_\_, "Phonology in Generative Grammar", Word 18.54-72 (1962). Reprinted in Fodor and Katz (1964).
- \_\_\_\_\_, "A descriptive convention for treating assimilation and dissimilation", Quarterly Progress Report 66, R.L.E., M.I.T., pp. 295-6 (1962a).
- \_\_\_\_\_, "The Great Vowel Shift" (forthcoming).
- \_\_\_\_\_, "Akan'e: the treatment of nondiffuse unstressed vowels in Southern Russian dialects (forthcoming a).

Halle, M. and N. Chomsky, The Sound Pattern of English (forthcoming).

\_\_\_\_\_, and T. Lightner, "Relative Chronology and the Synchronic Order of Rules" (forthcoming).

Harris, Z.S., Development of the Canaanite Dialects. New Haven: American Oriental Society (1939).

Hirt, H., Indogermanische Grammatik, I. Heidelberg: Winter (1921).

Hockett, C., A Course in Modern Linguistics. New York: Macmillan (1958)

\_\_\_\_\_, "Sound Change", Language 41.185-204 (1965).

Hoenigswald, H. M., Language Change and Linguistic Reconstruction. Chicago: Univ. of Chicago Press (1960).

\_\_\_\_\_, "Some Structural Effects of Grassmann's Law", paper presented at the annual winter meeting of the Linguistic Society of America (1964).

Hotzenköcherle, R., Die Mundart von Mutten (1935).

Jakobson, R., "Observations sur le classement phonologique des consonnes", Proceedings of the Third International Congress of Phonetic Sciences, Ghent (1939). Reprinted in Selected Writings, Vol. I.

\_\_\_\_\_, "Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves", TCLP II (1929). Reprinted in Selected Writings, Vol. I.

\_\_\_\_\_, "Über die phonologischen Sprachbünde", TCLP, IV (1931). Reprinted in Selected Writings, Vol. I.

\_\_\_\_\_, "Zur Struktur des russischen Verbums", Charisteria Guilelmo Mathesio quinquagenario oblata, Prague (1932). Reprinted in Vachek, J. (ed.), A Prague School Reader in Linguistics, Bloomington, Ind. (1964).

- Joos, M., Readings in Linguistics. New York:  
American Council of Learned Societies  
(1957).
- Jutz, L., Die alemannischen mundarten, Halle  
(Saale), (1931).
- Kattunen, L., Eestin kielen hännehistoria.  
Helsinki (1929).
- Klima, E., "Relatedness between Grammatical  
Systems", Language 40.1-20 (1964).
- \_\_\_\_\_, Studies in Diachronic Transformational  
Syntax. Harvard Dissertation (1965).
- Kuryłowicz, J., "La nature des procès dits  
'analogiques'", AL 5.15-37 (1949). Reprinted  
in Esquisses linguistiques, (1960).
- \_\_\_\_\_, The Inflectional Categories of  
Indo-European. Heidelberg: Winter (1964).
- Langendoen, D. T., "On a Restriction in the  
Application of Grassmann's Law in Greek"  
(forthcoming).
- Lejeune, M., Traité de phonétique greque. Paris (1947).
- Leskien, A., Die Deklination im Slavisch-  
Litauischen und Germanischen. Leipzig (1876).
- Lightner, T., "On the Phonology of old Church  
Slavonic Conjugation", International Journal  
of Slavic Linguistics and Poetics, Vol. 9 (1966).
- Lindgren, Kaj. B., Die Apokope des mhd. -e in  
seinen verschiedenen Funktionen. Ann. Ac.  
Sc. Fenn. B 78, 2 Helsinki (1953).
- Luick, K., Historische Grammatik der englischen  
Sprache (1921). Reprinted from the first  
edition: Oxford/Stuttgart (1964).
- Macdonell, A. A., Vedic Grammar (1910).
- McCawley, J., "Statement vs. Rule in Linguistic  
Description". Presented at the New York  
Linguistic Circle annual conference,  
March 13, (1965).

McCawley, J., The Phonological Component of a Grammar of Japanese. The Hague: Mouton & Co. (1966).

Meillet, A., Introduction à l'étude comparative des langues indo-européennes<sup>8</sup> (1937). Reprinted in Alabama Linguistic and Philosophical Series, Vol. 3.

\_\_\_\_\_, and J. Vendryes, Traité de grammaire comparée des langues classiques.<sup>3</sup> Paris (1963).

Moulton, W., "The Short Vowel Systems of Northern Switzerland", Word 16.155-182 (1960).

Niedermann, M., Historische Lautlehre des Lateinischen. Heidelberg (1953).

Paul, H., Prinzipien der Sprachgeschichte<sup>6</sup>, (1920).

Penzl, H., "Umlaut and Secondary Umlaut in Old High German", Language 25.223-40 (1949).

Postal, P., "On the Mentalistic Character of so-called 'Sound Change'" (forthcoming).

Rohlf, G., Historische Grammatik der italienischen Sprache, Band I (1949).

Rothe, W., Einführung in die historische Laut- und Formenlehre des Rumänischen. Halle (1957).

Sapir, E., "Notes on Judeo-German Phonology", The Jewish Quarterly Review, n.s. 6.231-66 (1915). Reprinted in Selected Writings of Edward Sapir.

de Saussure, F., "Sur un point de la phonétique des consonnes", Mémoires de la Société de Linguistique 6.246-57 (1889).

\_\_\_\_\_, Course in General Linguistics. Eds. C. Bally and A. Sechehaye, transl. W. Baskin. New York: Philosophical Library (1959).

Schane, S., French Phonology. Unpublished dissertation, M.I.T. (1965).

- Schirmunski, V. M., Deutsche Mundartkunde. Deutsche Akademie der Wissenschaften zu Berlin, Veröffentlichungen des Instituts für deutsche Sprache und Literatur, Band 25. Berlin (1962).
- Schmidt, J., Die Verwandtschaftsverhältnisse der indogermanischen Sprachen. Weimar (1872).
- Schuchardt, H., Über die Lautgesetze. Gegen die Jung-grammatiker. Berlin: Oppenheim (1885). Page references are to the reprinting in L. Spitzer, (ed.) Hugo Schuchardt-Brevier. Halle: Niemeyer (1922).
- Schwyzer, E., Griechische Grammatik, I. München (1939).
- Sommer, I., Handbuch der lateinischen Laut- und Formenlehre. Heidelberg (1948).
- Sommerfelt, A., "The Development of Quantity as Evidence of Western European Linguistic Interdependence", in Selected Writings. Mouton: The Hague (1963).
- Spence, N.C.W., "Quantity and Quality in the Vowel System of Vulgar Latin", Word 21.1-18 (1965).
- Streitberg, W., Urgermanische Grammatik<sup>3</sup>. Heidelberg: Winter (1963).
- Streitberg, W., "Zur Geschichte der Sprachwissenschaft", Indogermanische Forschungen 35.182-96 (1915).
- Thieme, P., Review of Albert Carnoy, Dictionnaire étymologique du proto-indo-européen. Language 34.510-5 (1958).
- Thumb, A. and R. Hauschild, Handbuch das Sanskrit, I. Heidelberg (1958).
- Twaddell, W. F., "A note on Old High German Umlaut", Momatshefte für deutschen Unterricht 30.177-81 (1938). Reprinted in Joos (1957).
- Vaillant, A., Grammaire Comparée des Langues Slaves, Tome I. Lyon/Paris (1950).

- Vogt, H., "Les occlusives en arménien", Norsk tidskrift for Sprogvidanskap, Vol. 18 (1958).

Wackernagel, J., Altindische Grammatik, I. Halle (1896).

Wanner, G., Die Mundart des Kantons Schaffhausen. Vol. 20 of A. Bachmann (ed.) Beiträge zur Schweizerdeutschen Grammatik (1941).

Watkins, C., "Problems in Old Irish Phonology" (forthcoming).

Weinreich, U., "Four Riddles in Bilingual Dialectology", American Contributions to the Fifth International Congress of Slavists, 335-58 (1963).

Wheeler, B. I., Der Griechische Nominalakzent. Strassburg (1885).

Whitney, W. D., Sanskrit Grammar<sup>2</sup> (1889).

Zwickly, A., Topics in Sanskrit Phonology. M.I.T. Dissertation (1965).

Manczak, W., "Origine de l'apophonie e/o en indo-européen", Lingua 9.277-87 (1960).



3 1262 04592 7950

765

**U. OF F. LIBRARIES**